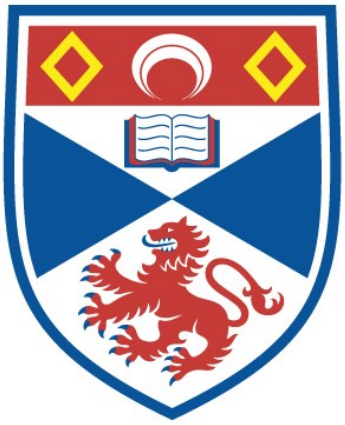
STEREOCHEMICAL AND MECHANISTIC STUDIES ON THE ASPARTIC PROTEASES

Paul Charles David Hawkins

A Thesis Submitted for the Degree of PhD
at the
University of St Andrews



1993

STEREOCHEMICAL AND MECHANISTIC STUDIES ON THE ASPARTIC PROTEASES

a thesis presented by

Paul Charles David Hawkins

to the

UNIVERSITY OF ST. ANDREWS

in application for

THE DEGREE OF DOCTOR OF PHILOSOPHY

St Andrews

March 1993

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

DECLARATION

I, Paul Charles David Hawkins, hereby certify that this thesis has been composed by myself, that it is a record of my own work, and that it has not been accepted in partial or complete fulfiment of any other degree or professional qualification.

Signed Paul Hawkins Date 1/31/'93

I was admitted to the Faculty of Science of the University of St. Andrews under Ordinance General No. 12 on 1st Oct 1990 and as a candidate for the degree of Ph.D. on 1st Oct 1990

Signed Paul Hawkins Date 1/31/'93

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate to the degree of Ph.D.

Signature of supervisor ....

Date 1st March '93

Copyright

To

my parents and my sister

Acknowledgements

I would like to thank my friend and supervisor, Prof. D. Gani, for his continual support and enthusiasm throughout out the entirety of the project and for giving me my head when all around me were losing theirs.

Thanks are due to all the members of the Gani group down the ages for their patience and help. I will first thank my partners in crime on the HIV project, Nick Camp and David "Merry" Perrey for admirable assistance and amusement. Particular thanks are due to Dr. Mahmoud Akhtar for his help and fine food, Dr. Nigel Botting, Amit "M.C." Mehrotra for his hatchet job on parts of the thesis and to Basil for his friendship. All of this work would have been much more difficult without the patience and helpfulness of Miss Stacey Low, for which I am forever in her debt.

Thanks are also due to Dr. Ron Hay and the denizens of Lab 28, especially Dr. Angela Dawson, Dr. James Matthews, Dr. Paul "Shadowy" Szawlowski and Mr. Alan Monaghan for all their help with the molecular biology and for making me feel so welcome. My gratitude goes to Dr. Martin Ryan of the AFRC's AVRI at Pirbright for the generous gift of a number of vectors and for his valuable advice.

I must acknowledge the assistance of Dr. Derek Kinchington, St. Bartholomew's Hospital, London for carrying out the *in vivo* tests on our HIV protease inhibitors, Dr. I. Jones, NERC Institiute of Virology for the gift of plasmid pRP14, the MRC ADP for supplies of HIV-1 protease and substrate and Mr. T. Hanke for the gift of baculovirus.

I must thank my parents, without whom none of this would have been possible, for their constant encouragement throughout.

Finally, thanks to the SERC for financial support.

ABSTRACT

A series of experiments, designed to investigate two mutually incompatible theories of the catalysis carried out by pepsin, the archetypal aspartic protease, were undertaken. Mechanism-activated active-site probes, based on acyl hydrazides, were synthesised, but could not be shown to inactivate pepsin. Experiments designed to trap a covalent intermediate in pepsin catalysis were also carried out, but did not provide any evidence for such an intermediate.

A number of methyl hydrogen 1-aminoalkyl phosphonates have been synthesised. They were used, by co-workers in the group, in the synthesis of phosphonamidate-containing penta- and hexapeptide-based inhibitors for the aspartic protease from HIV-1. These compounds were found to be inhibitors in both *in vitro* and *in vivo* assays. The best inhibitors had IC_{50} values in the low micromolar range. Molecular modelling was used to develop a model for the interaction of these compounds with the active site of the protease. This model was used to rationalise some of the results obtained from the inhibitors.

Attempts were made to clone, overexpress and purify the protease from *E. coli*. The protease was purified to homogeneity but no activity could be observed. Various attempts to obtain activity were unsuccessful.

Contents

List of Figures

List of Tables

Abbreviation	Meaning
Ac	N-acetyl
AMV	avian myoblastosis virus
APC	antigen presenting cell
ATP	adenosine 5' triphosphate
BCIG	5-bromo-4-chloro-3-indolyl-β-D-galacto-pyranoside
Boc	tertiarybutoxy carbonyl
n-BuLi	*normal*-butyllithium
bp	base pairs
BSA	bovine serum albumin
Cbz	carbobenzyloxy
Da	dalton
DMF	dimethylformamide
DMSO	dimethyl sulphoxide
DNA	deoxyribonucleic acid
DTT	D,L dithiothreitol
E. coli	Escherichia coli
E.C.	Enzyme Catalogue
EDTA	ethylene diaminetetraacetic acid
ELISA	enzyme-linked immunosorbent assay
Fmoc	9-fluorenylmethoxycarbonyl
GST	glutathione-S-transferase
HIV	human immunodeficiency virus
HLA	human leukocyte antigen
HOBT	hydroxy benzotriazole

IC_{50}	concentration required to reduce enzyme activity by 50%
IPTG	isopropyl ß-D-thioglucopyranoside
K_i	enzyme inhibition constant
K_M	Michaelis-Menten constant
LB	Luria-Bertani
LDA	lithium di*iso*propyl amide
mCPBA	*meta*-chloroperbenzoic acid
MCS	multiple cloning site
MHC	major histocompatibility complex
MOPS	3-[N-morpholino]propanesulfonic acid
NMR	nuclear magnetic resonance
Phe(NO_2)	*para*-nitrophenylalanine
OD	optical density
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PEG	polyethylene glycol
pfu	plaque forming units
PMSF	phenylmethylsulfonyl fluoride
RF	replicative form
r.m.s.	root mean square
r.p.m.	revolutions per minute
RNA	ribonucleic acid
RSV	Rous sarcoma virus
SDM	site directed mutagenesis
SDS	sodium dodecyl sulfate
TCID	tissue culture infective dose

TCR	T-cell receptor
TEN	Tris / EDTA/ sodium chloride
THF	tetrahydrofuran
tlc	thin layer chromatography
Tris	tris(hydroxymethyl)aminomethane
UV	ultraviolet

The one and three letter codes for the amino acids

Amino acid	Three letter code	Single letter code
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cyclohexylalanine	Cha	-
Cysteine	Cys	C
Glutamic acid	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

CHAPTER ONE

INTRODUCTION TO THE ASPARTIC PROTEASES AND HIV-1

1.0 Introduction: The aspartic proteases

The aspartic proteases are one of the four known classes of proteolytic enzymes that act on interior bonds of proteins (endopeptidases); the other classes are serine, cysteine and metallo proteases. The aspartic proteases are so named for the presence of two absolutely conserved aspartate residues in the active site (see below). The optimal pH of many of these enzymes lies in the range 1-3 and consequently they have been previously designated the acid proteases. However, since the discovery of several members of this class that have pH optima near 7 the term aspartic protease has been adopted. The aspartic proteases are ubiquitous, occurring in most phyla, *e.g.* in viruses (HIV-1 and 2, RSV and AMV), bacteria (*B. subtillus*), protozoa (*Tetrahymena*), fungi (*Endothia parasitica*), higher plants (*Lotus*) and vertebrates. They are involved in a wide range of functions *in vivo*, including digestion of dietary protein (pepsins and chymosin), homeostatic regulation (renin) and lysosomal protein degradation (cathepsins D and E). The fungal proteases are believed to be involved in sporulation and the retroviral proteases cleave viral polyproteins during maturation of the virus. The aspartic proteases are also involved in a number of pathological conditions; pepsin in gastric ulcers, renin in hypertension and cathepsins D and E in muscular dystrophy and neoplastic diseases. The aspartic proteases have also been extensively used in food processing, for the production of cheese and fermented foods from soya beans.

All of the aspartic proteases from these widely differing sources share a high degree of structural and sequence homology. They usually have acidic pH optima and are inhibited by carboxyl group directed inhibitors and the fungal metabolite pepstatin A, which is the general aspartic protease inhibitor.[1]

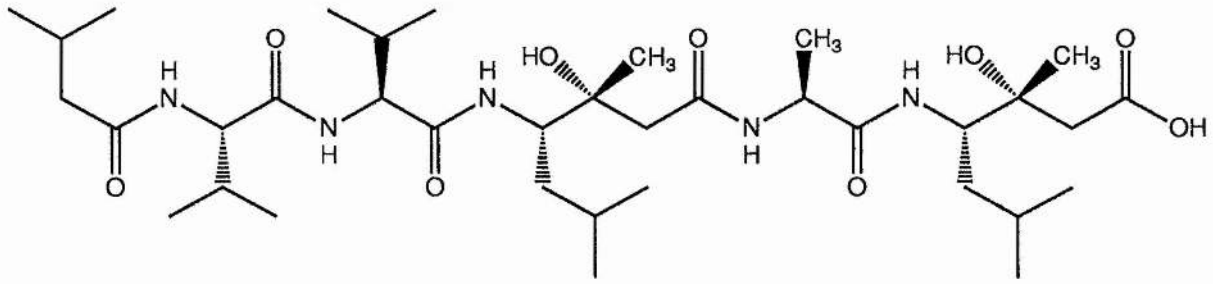


Figure 1.0.1: Pepstatin A

1.1 Protease action

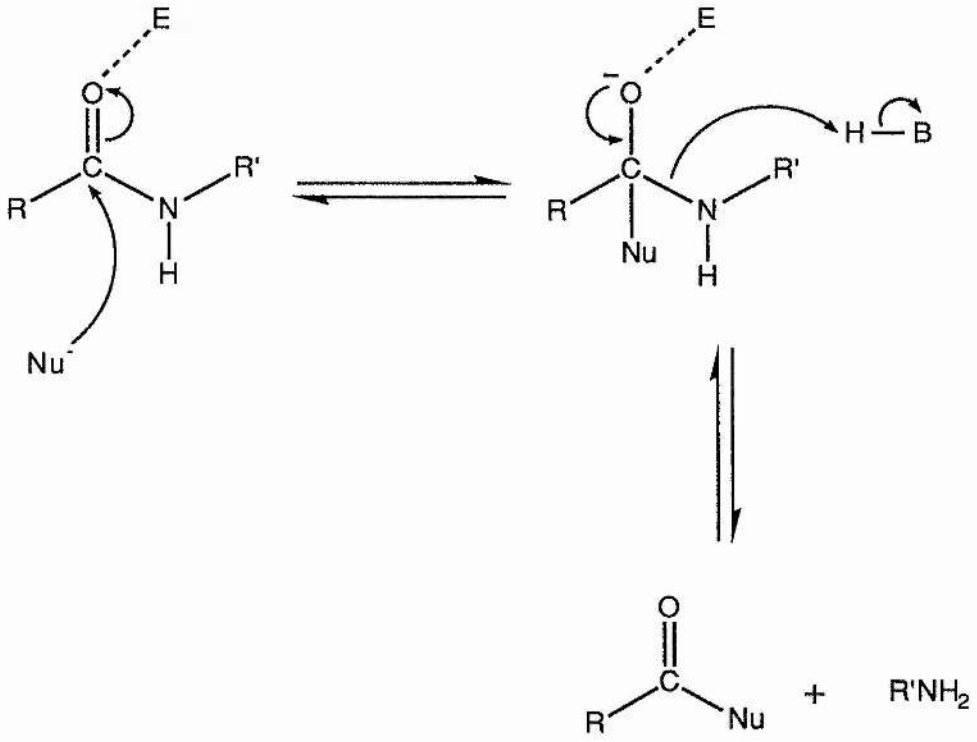
All proteolytic enzymes carry out the same superficially simple reaction, the addition of water across an amide bond. They have, however, all adopted different methods to catalyse this difficult reaction, whilst providing the same factors to achieve catalysis. The factors that must be provided by a protease to achieve peptide bond cleavage are;

i) a Lewis or Brønsted acid (E) to polarise the scissile carbonyl,

ii) a nucleophile (Nu) to attack the scissile carbonyl and

iii) a proton donor (B-H) to protonate the departing amine[2] (see Scheme 1.1.1 below).

Proteases need to involve all these different factors in the catalysis to overcome the enormous stability of the peptide bond, usually ascribed to its large resonance energy [3] of around 17-18 kcal mol^{-1}. Recently, the concept of stabilisation of amide bonds by resonance has been the subject of some debate.[4]



Scheme 1.1.1: Generalised protease mechanism
(Nu is water for the aspartic and metallo proteases.)

The serine and cysteine proteases utilise covalent acyl intermediates in the cleavage, acylating a nucleophilic hydroxyl or thiol side chain in the active site. In contrast the action of the metallo and aspartic proteases involves a nucleophilic water molecule and thus these systems are believed to operate by non-covalent mechanisms (see below). In the aspartic proteases the residues acting as the Lewis or Brønsted acid and the proton donor are not known with certainty.

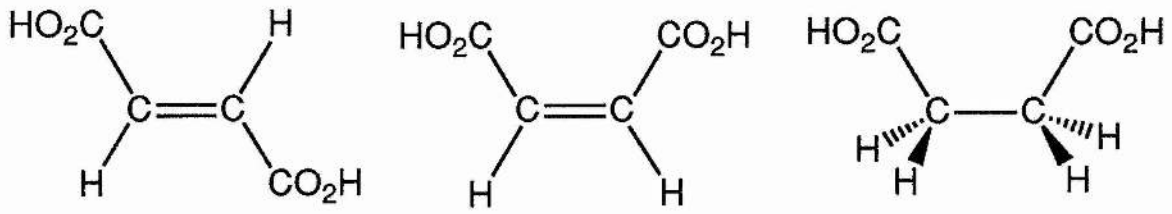
1.2 Molecular properties of aspartic proteases

1.2.1 Physico-chemical properties

The aspartic proteases are globular proteins with molecular weights in the range of 31-40,000 Da, that of porcine pepsin A is 34,644 Da, calculated from its primary sequence.[5] The aspartic proteases are highly acidic proteins, with a large preponderance of acidic over basic residues (40:3 found in the sequence of porcine pepsin A).[5] This results in a very low pI for these proteins, in the range 2-3, which is consistent with their low pH optimum.

The pH-rate profile of the aspartic proteases has been extensively investigated and studies of most systems, including pepsin, have revealed a low pK_a and a high pK_a. In pepsin the pK_a values were found to be 1.2 and 4.5.[6] This led the early investigators to propose that two carboxyl groups, one ionised and one not, were crucial to the catalysis.[7] The origin of this large difference in the pK_a's of the two carboxyls has been considered frequently. It has often been compared to the pK_a difference in 1,2 dicarboxylic acids[8] *e.g.* the pK_a's for maleic acid are 1.83 and 6.07.[9] The mandatorily *cis* arrangement of the carboxyl groups in maleic acid permits the formation of very strong intramolecular hydrogen bonds. The hydrogen bond between the carboxylate and the hydroxyl group of the undissociated carboxyl will greatly stabilise the mono-anion. However, the dianion will be destabilised as the negatively charged carboxylates will be forced to be close and co-planar. These two effects combine to produce the two widely separated pK_a's of maleic acid, as the first ionisation is promoted and the second disfavoured. The two pK_a's of succinic acid are, by contrast, 4.16 and 5.61 and those of fumaric acid are 3.03 and 4.44. Thus, the active site of the aspartic proteases may constrain its two carboxylate side chains to be *cis* and coplanar more than they are constrained in succinic acid, but less so than

in maleic acid.



Fumaric acid Maleic acid Succinic acid

Figure 1.2.1: Dicarboxylic acids

In order to probe the possible involvement of carboxyl groups in the catalysis by aspartic proteases chemical labelling experiments using various carboxyl-directed agents were undertaken. Among the reagents that have been used are trimethyloxonium fluoroborate,[10] diazoacetyl amino acid methyl esters,[11] 1-(*p*-nitrophenoxy)-2,3-propoxide (EPNP)[12] and *p*-bromophenacyl bromide.[13] All of these react with aspartic proteases to label at least one carboxyl group. Labelling of either one of the two carboxyl groups results in inactivation of the enzyme, as was shown by Delpierre and Fruton[14] using ^{14}C-labelled L-1-diazo-4-phenyl-3-tosylamidobutan-2-one (Tos-L-Phe-CHN_2). The rate of incorporation of ^{14}C label into the protein parallelled the loss of proteolytic activity against either peptide or protein substrates. This work was in accord with the obligate involvement of 2 carboxyl groups in the mechanism of the aspartic proteases. The origin of the active site carboxylates was established as aspartate residues by labelling with hydroxylamine and performing a Lossen rearrangement to give 2,3 diaminopropanoic acid.[15]

Experiments with EPNP showed incorporation of 2 molecules per molecule of pepsin A, one at Asp-32 and the other at the same Asp as was labelled by the diazo reagents.[3a,b] This residue is Asp-215 in the sequence of pepsin.[16]

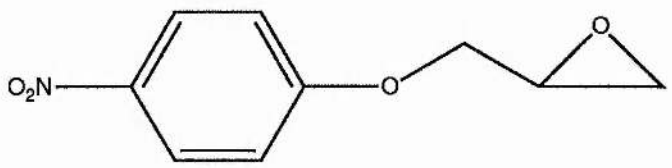


Figure 1.2.2: 1-(p-nitrophenoxy)-2,3-propoxide, EPNP

More recently this epoxide-based approach has been used to develop irreversible active site directed inhibitors for the aspartic protease from HIV-1[17] (see Section 1.6). It has been shown by electrospray mass spectroscopy and X-ray crystallography[17b] that the epoxide becomes covalently bound to one of the active site aspartates of the protease from HIV-1, as EPNP does with the monomeric proteases.[3]

The lower pK_a in pepsin was assigned to Asp-215 due to its selective reaction with diazo compounds such as N-diazoacetyl-Phe-3-phenyl 2,3-^{3}H propylamide.[18] However, this assignment is not unambiguous as free energy perturbation calculations have shown that the protonation states of the two aspartates are affected by the nature of the compound bound at the active site.[19] These uncertainties may explain some contradictory results obtained with other active-site labelling reagents that were used to locate the ionised aspartate. The different reagents may induce the two aspartates to adopt different protonation states, therefore giving label incorporated at both of the carboxyl groups, dependent upon the group used to carry out the labelling.

1.2.2 Structural and sequence properties

The fungal and mammalian aspartic proteases all adopt very similar tertiary folding patterns. They are globular proteins with the overall structure consisting of 2 lobes, made up of the N- and C-terminal parts of the sequence. The active site lies in a deep cleft at the junction between the two lobes. The active site appears to be approximately 25 Å long in the X-ray crystal structures (see Section 1.4), which is consistent with the ability of the aspartic proteases to bind substrates up to 7 amino acids in length.[20] The active site is sealed off from solvent upon substrate binding by a structure known as the flap (residues 71-83 in pepsin). After catalysis the flap opens to allow products to leave and new substrate to bind. The flap is clearly of pivotal importance in the productive binding of substrates as it sequence is highly conserved.[21]

The aspartic proteases consist mostly of ß-sheet. Indeed pepsin is around 44% ß sheet.[22] An extensive six stranded antiparallel ß-sheet forms the core of the enzyme, three strands being contributed by each of the N- and C-terminal domains. This

dominance of ß-sheet structure has led to the aspartic proteases being classified as "all ß" proteins.[23] The high proportion of ß-sheet structure and the low amount of α-helical structure is also seen in circular dichroism studies of pepsin in solution.[24]

A large number of aspartic proteases from widely differing sources have been sequenced[21] and the sequences show significant homology in some areas. The sequences around the 2 active site aspartates, 32 and 215 in porcine pepsin, are highly conserved, as would be expected. Homology is also observed around Tyr-75 in the flap,[21] indicating that this region too is crucial for efficient catalysis. This conservation of sequence may indicate evolution from a common ancestor.

The D-T-G-(S/T) motif is considered to be diagnostic for the aspartic proteases, and has been used to identify aspartic proteases from their sequence, for example the HIV-1 protease (see Section 1.6). This motif is important in maintaining the structure of the active site through hydrogen bonding (see Section 1.4). The strict conservation of this motif is shown in the sequence alignments in Table 1.1.

Table 1.1 *Partial sequence alignment for some aspartic proteases.*

Enzyme	Sequence around residue:		
	32	75	215
Endothiapepsin	**D**TGSSD	**Y**GDG	**D**TGTT
Penicillopepsin	**D**TGSSD	**Y**GDG	**D**TGTT
Rat renin	**D**TGSAN	**Y**GSG	**D**TGTS
Porcine pepsin	**D**TGSSN	**Y**GTG	**D**TGTS
HIV PR	**D**TGADD	--	--

1.2.3 Enzymology

As mentioned above the aspartic proteases have a very low pH optimum. The reasons for this have been examined in some detail (see Section 1.2.1). The hydrogen bonding network around the catalytic aspartates (see Section 1.4) is believed to greatly stabilise the mono-anionic form of the enzyme. This results in the very low first pK_a discussed above (see Section 1.2.1). The high pH optimum of human renin (around 5) was believed to be due to it having an alanine residue at position 218. This residue cannot hydrogen bond to aspartate 215 and thereby stabilise a negative charge. However, as shown above (Table 1.1) rat renin, which has a pH optimum almost identical to that of human renin, has a threonine residue at position 218. Thus, the hydrogen bond from this residue to Asp-215 appears to have little influence on the pH optimum for the enzyme. It has been proposed that the pH optimum of renin is influenced greatly by its substrate, angiotensinogen[25] and is not determined by the active site sequence. This is supported by the work of Tang[26] on mutagenesis of pepsin and rhizopuspepsin active sites. Mutation of the serine/threonine residues at positions 35 and 218 in rhizopuspepsin to alanine had no effect on the pH-rate profile for the enzyme, though the catalytic efficiency was decreased around ten fold. It was therefore postulated that the hydrogen bond only contributes to active site rigidity and not to the pK_a of the apartates or the pH optimum of the enzyme. A decrease in active site rigidity could lead to a randomisation of the position of the active site water molecule and a consequent decrease in catalytic efficiency.

A similar study was carried out on the aspartic protease from HIV-1,[27] replacing Ala-28, which is in the equivalent position to Ser/Thr-218 in the monomeric proteases, with serine. It was found that the mutant enzyme showed a small lowering of the optimum pH, but not as large as would be expected if the mutation were to confer similar pH properties to the monomeric proteases. The mutant also showed a decrease in k_{cat}, in contrast to the results from the monomeric proteases. This may be due to the HIV-1 protease requiring a more flexible active site to accomodate its wide range of substrates. The hydroxyl group of Ser/Thr-218 therefore probably does not interact with the active site aspartates as expected in either the monomeric or dimeric aspartic proteases. The origin of the very low first pK_a in most of the monomeric proteases remains, therefore, unclear. It may well derive from the coplanar conformation imposed upon the two aspartates by the active site, as

discussed above (Section 1.2.1).

Most of the aspartic proteases have little specificity in their requirements for the amino acids flanking the scissile bond *i.e.* the S_1 and S_1' subsites (see Figure 1.2.3), in the nomenclature of Schecter and Berger,[28] are not very stringent.

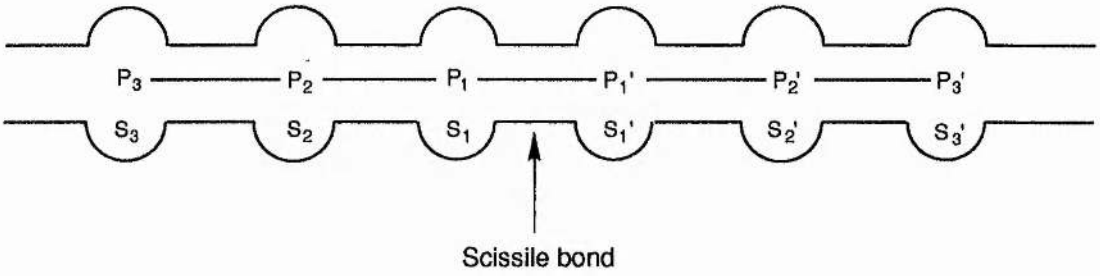


Figure 1.2.3: Substrate and subsite nomenclature according to Schecter and Berger[28]

The best characterised exception is renin, which has only 1 natural cleavage site, the Leu 10-Val 11 bond in angiotensinogen. This specificity difference is consistent with the respective roles of the more indiscriminate digestive protease such as pepsin and the exquisitely tuned specificity of a regulatory protease like renin.

A prominent feature of aspartic protease catalysis is the observation that k_{cat} for peptide cleavage increases with increasing length of the substrate while K_M remains approximately constant.[29] Thus, the extra binding energy available from binding larger substrates is used to lower the activation energy for the hydrolysis and is not expressed in increased strength of binding. The reduction in activation energy results in an increase in k_{cat}, but just how this is achieved has been the subject of some debate. It has been shown that the rate-determining step in aspartic protease catalysis is the loss of products from the active site[30] and not the cleavage of the scissile peptide bond. Thus, differences in k_{cat} between different substrates arise from differences in the affinity of the active site for the products of cleavage and not from differences in the stability of the scissile bond.

All aspartic proteases have a significant electrophilic component in their catalysis of peptide bond hydrolysis. This electrophilicity is suggested by the observation that the aspartic proteases, in common with electrophilic systems such as H^+, catalyse the hydrolysis of esters and amides at approximately equivalent rates. This stands in contrast to nucleophilic systems such as OH^- and the serine proteases which

hydrolyse esters much more quickly than amides. Recent *ab initio* calculations have shown that esters are much more likely to be hydrolysed *via* an anionic intermediate whereas amides are more likely to be hydrolysed by a neutral intermediate.[31] Thus, the aspartic proteases may well catalyse peptide bond hydrolysis *via* a neutral intermediate (see Section 1.4).

The aspartic proteases are some of the most efficient proteases in Nature, as shown in Table 1.2 below:

Table 1.2 *Kinetic constants for representative proteases of each class.*

Enzyme	Substrate	k_{cat}/K_M ($M^{-1}s^{-1}$)
Chymotrypsin[32] (serine)	AcPAPF*AAANH_2	4.5×10^4
Papain[33] (cysteine)	ZGVE*LG	3×10^3
Pepsin[32] (aspartic)	ZGAF*FOP_4P	4.1×10^7
CarboxypeptidaseA[33] (metallo)	BzGGG*LOPh	1.5×10^6

The scissile bond is represented by *. OP_4P is 4-pyridyl propoxy.

Apart from hydrolysis the aspartic proteases exhibit two other activities against peptide substrates, condensation and transpeptidation. Condensation, the direct joining of two peptides, one of which need not necessarily be a substrate, to form a larger peptide is the basis of action of the activator peptides. Transpeptidation, the joining of a fragment of a substrate peptide to another substrate peptide, is discussed in detail in Sections 1.3.1.1 and 1.3.1.2.

1.3 The mechanism of the aspartic proteases

The large amount of active site sequence homology and sensitivity to the same inhibitors among the aspartic proteases indicates that they all the act by the same mechanism. Any enzyme mechanism has three components;[34]

i) Kinetic - which addresses the binding/debinding order of substrates and products.

ii) Chemical - which considers the chemical species (intermediates and transition states) involved and the free energy profile for the reaction.

iii) Structural - which attempts to show how the enzyme actually lowers the activation energy for the reaction catalysed.

Historically, investigations into the mechanism of the aspartic proteases followed roughly this path, the first proposed mechanisms being based on kinetic studies and product analysis, then on chemical studies and then on X-ray crystal structures.

The nature of aspartic protease catalysis has been the subject of much experimentation and many hypotheses. The two active site aspartates must be involved, (see above) but their exact role has been unclear, with two classes of mechanism having been proposed. The covalent mechanisms postulate a covalent intermediate involving part of the substrates and one of the active site aspartates. The general base mechanisms propose that one of the aspartate residues forms the attacking nucleophile from water bound at the active site. Thus, either one of the active site aspartate side chains or an active site water molecule is the nucleophile.

1.3.1 The covalent mechanisms

The covalent mechanisms were the first mechanisms put forward to account for the different types of activity of the aspartic proteases. They seemed to economically account for the three different activities exhibited by the aspartic proteases acting on peptide substrates; hydrolysis, condensation and transpeptidation. Transpeptidation, the formation of larger peptides from smaller ones by adding a portion of a substrate to itself or another substrate peptide, is a unique property of most of the aspartic proteases. There are two types of transpeptidation activity, amino and acyl, and they shall be considered in turn. Transpeptidation of either sort is only observed with poor substrates and is only seen at pH's close to neutrality (4-6).[35]

1.3.1.1. The amino enzyme mechanism

The existence of a covalent amino enzyme intermediate was proposed after amino transpeptidation was observed in studies[35] of pepsin acting on di- and tripeptides. The carboxy terminal amino acid of the substrate was observed to be transferred to a suitable peptidic acceptor, either another peptide or another molecule of substrate. This became known as amino transpeptidation. Amino transpeptidation was then observed in several similar systems.[36] The net effect of amino transpeptidation is to exchange the amino terminus of the original peptide substrate (H_2N-X-CO) for the incoming peptide substrate (Z-CO_2H), as shown below.

$$Z\text{-}CO_2H + H_2N\text{-}X\text{-}CONH\text{-}Y\text{-}CO_2H \rightleftharpoons Z\text{-}CONH\text{-}Y\text{-}CO_2H + H_2N\text{-}X\text{-}CO_2H$$

Scheme 1.3.1: Amino transpeptidation
where Z may be substrate or a different peptide

The "amino enzyme" mechanism was developed to account for the observation of amino transpeptidation products.

$$\mathbf{E}\text{-}CO_2H + H_2NX\text{-}CONHY\text{-}CO_2H \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \mathbf{E}\text{-}CO_2H[X\text{-}CONH\text{-}YCO_2H] \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} \mathbf{E}\text{-}CONH\text{-}YCO_2H + H_2N\text{-}X\text{-}CO_2H$$

$$\mathbf{E}\text{-}CO\text{-}NH\text{-}Y\text{-}CO_2H + H_2O \underset{k_{-3}}{\overset{k_3}{\rightleftharpoons}} \mathbf{E}\text{-}CO_2H + H_2N\text{-}Y\text{-}CO_2H$$

Scheme 1.3.2: Amino enzyme hydrolysis
where **E** represents the enzyme

The amino enzyme mechanism postulated the formation of a covalent bond between the enzyme and the carboxy portion of the peptide substrate (NH-Y-CO_2H), to form a so-called 'amino enzyme' intermediate. Two different fates could befall this intermediate, **E**-CO-HN-Y, either hydrolysis to yield H_2N-Y-CO_2H or reaction with another peptide or amino acid acceptor to give the amino transpeptidation product.

another peptide or amino acid acceptor to give the amino transpeptidation product.

In the scheme above k_1 is much greater than k_2, and k_{-1} is much greater than k_{-2} *i.e.* a fast reversible formation of the Michaelis complex is followed by relatively slow chemical step(s) and/or conformational changes (see below). The net result is the production of a long peptide from shorter ones.

The acceptor in amino transpeptidation is the carboxyl group of the substrate or other peptide as Neumann *et al.*[37] showed that transpeptidation did not occur if there was no free carboxyl group available. Silver and James[38] have shown that the carboxylate anion is required for amino transpeptidation. The inability to trap the covalent intermediate[35] showed that if it exists, it must occur on the reaction pathway after the rate-determining step. If it occurred before the rate-limiting step then the concentration of the intermediate would build up and could therefore be trapped. Also, no burst release of product (H_2N-X-CO_2H) is observed, which would be expected if the covalent intermediate occurred before the rate-limiting step. The mechanism outlined schematically above for amino transpeptidation makes the prediction of ordered product release, N-terminal portion of the substrate first. The acceptor peptide and the carboxyl portion of the substrate will not be bound at the same time as they are assumed to occupy the same portion of the active site.

This was supported by Fruton *et al.*[39], who showed that incubation of pepsin with Cbz-Tyr-Tyr, Cbz-^{14}C-L-Tyr and L-Tyr resulted in ^{14}C-label incorporation into Cbz-Tyr-Tyr at the N-terminal residue. Incubation with ^{14}C-L-Tyr and Cbz-L-Tyr, however, resulted in no incorporation of radioactivity. Therefore, Cbz-L-Tyr is released from the active site and the C-terminal L-Tyr is retained long enough for exogenous labelled Cbz-L-Tyr to enter the active site and combine with it. This is consistent with the kinetics predicted for a mechanism involving an amino intermediate. Ordered product release was further supported by the work of Knowles[40] in which it was shown that the hydrolysis of Ac-Phe-Phe-OMe is inhibited non-competitively by Ac-L-Phe and competitively by L-Phe-OMe. This is in accord with a kinetic mechanism in which Ac-L-Phe is released first.

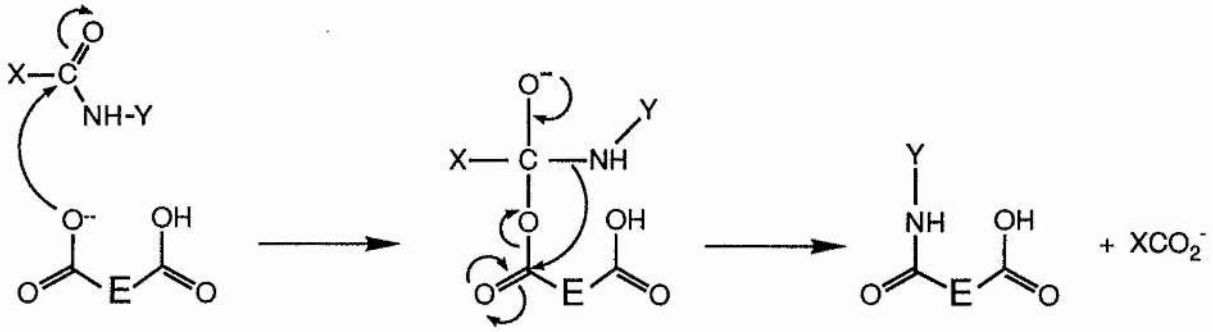
The kinetics of product release show a pH dependency. At pH 2.1, when the carboxyl group is mostly undissociated, Ac-L-Phe is an uncompetitive inhibitor. However, at pH 4.3, when the carboxyl is mostly dissociated, Ac-L-Phe approximates to a competitive inhibitor. This indicates that the Ac-L-Phe leaves as an anion from

the hydrolysis and enters as an anion during transpeptidation. This supports the work of Silver and James[38] on the obligate involvement of the carboxylate anion in amino transpeptidation.

The "amino enzyme" formulation received additional support from isotope exchange studies of Shkarenkova *et. al.*[41] The failure to observe exchange of free amino acids with enzyme-bound amino acid intermediates by Sharon *et al.*[42] is consistent with a covalently bound intermediate. Sharon *et al.*[42] also found that ^{18}O-label is exchanged between L-Phe (and not D-Phe) and $H_2{}^{18}O$ only in the presence of pepsin (*i.e.* the exchange is enzyme catalysed, see also[43]).

Thus the amino mechanism seemed to be well supported experimentally. However, the work of Knowles was re-examined[44] and it was found that Ac-L-Phe inhibits the cleavage of Ac-L-Phe-L-Tyr partially competitively and partially non-competitively. By studying this effect over a range of pH values it was concluded that the order of the dissociation of the ternary enzyme-product complex is pH dependent and thus offers no real information on the order of product release. A similar study[45] was carried out on Ac-L-Phe-L-Tyr-OEt and at pH 4.7 ordered release of Ac-L-Phe first and L-Tyr-OEt next was observed *i.e.* an ordered-off mechanism, in accord with an amino enzyme mechanism. However, this effect was not observed at pH less than 4 as the kinetic mechanism changes to a random or rapid equilibrium type at around pH 3.5. The authors suggest that this is due to a group of pK_a around 4 being involved in binding the hydrolysis product L-Tyr-OEt at the active site and this binding becomes weakened at pH less than 4, allowing the L-Tyr-OEt to leave.

There were several other difficulties with the amino enzyme hypothesis. The chemical mechanism for the formation of the 'amino enzyme' was particularly problematic, the most plausible theory involving a four-centre exchange reaction (as shown in Scheme 1.3.3). This mechanism was first proposed by Knowles,[46] with a variant later proposed by Clement.[47]



Scheme 1.3.3: Possible amino enzyme mechanism

where E represents the enzyme.

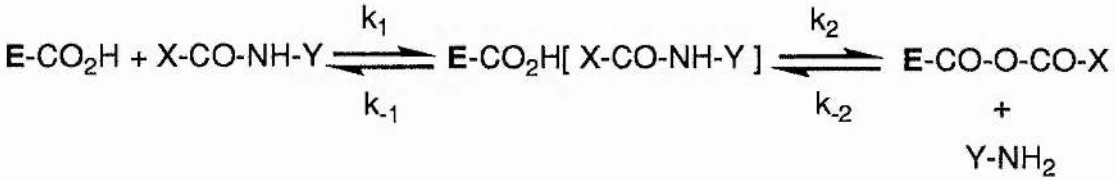
The hypothesis of an intermediate of the type '**E**-NH-Y' as the sole intermediate on the hydrolysis pathway was shown to be untenable by Silver *et al.*[48] Their work showed that small changes in the nature of X in X-CO-NH-Y, *i.e.* the N-terminal portion of the substrate, produced very large changes in the rates of amino transpeptidation and hydrolysis. This is not consistent with an intermediate of the type E-NH-Y being involved in catalysis. There were doubts cast by Silver and Stoddard[49] upon the interpretation of the product inhibition data of Knowles[40] that had indicated the involvement of an amino-type intermediate. Another problem is the chemical mechanism of the reaction between the carboxylate form of the acceptor peptide and this intermediate to form the transpeptidation product. It is not trivial to develop a chemically plausible mechanism for this reaction. Hunkapiller and Richards[50] first suggested that the amino enzyme intermediate may not involve a covalent bond, but did not propose a non-covalent mechanism for the hydrolysis. They instead proposed an acyl intermediate of the type discussed below. As the difficulties with the amino enzyme mechanism grew and other transpeptidation activity was discovered attention was turned to a different mechanism.

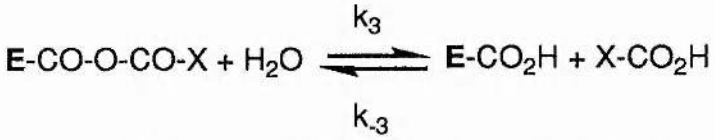
1.3.1.2 The acyl enzyme mechanism

The observation of acyl transpeptidation[51] indicated that the 'acyl enzyme' should exist, just as amino transpeptidation had suggested the existence of an 'amino enzyme' intermediate. Acyl transpeptidation involves the transfer of the amino portion of the substrate to a suitable peptide acceptor.

Scheme 1.3.4: Acyl transpeptidation

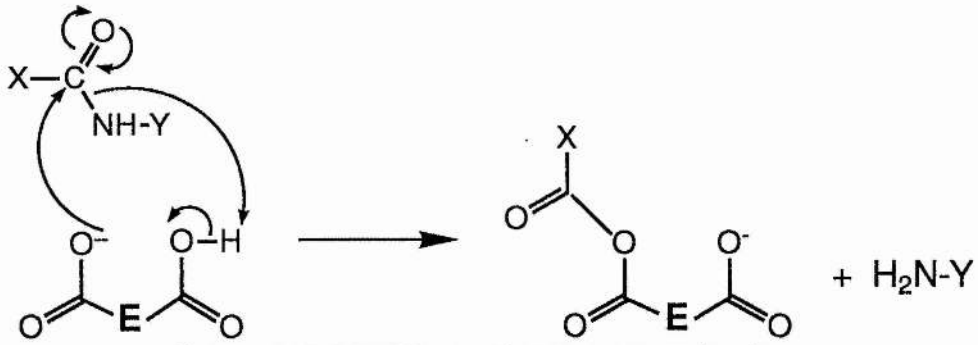
An acyl intermediate was proposed as the central intermediate in acyl transpeptidation. This hypothesis was supported by the work of Akhtar *et al.*[52] Their results from trapping experiments using tritiated methanol and their observation of a covalently bound intermediate in the hydrolysis of Cbz-Tyr-[^{3}H]Tyr were consistent with an acyl intermediate. However, these experiments proved to be irreproducible and were later retracted.[53] The overall pathway of hydrolysis involving a covalent acyl intermediate (**E**-CO-O-CO-X) is outlined below in Scheme 1.3.5.





Scheme 1.3.5: Acyl enzyme hydrolysis
where **E** represents the enzyme

Acyl transpeptidation occurs, instead of hydrolysis, when the acyl intermediate is intercepted by a free amino group, from either the substrate or another peptide. The mechanism of formation of the acyl intermediate (see Scheme 1.3.6 below) is much simpler than the mechanism for the amino intermediate formation.



Scheme 1.3.6: The acyl enzyme mechanism
where E represents the enzyme

Hunkapiller and Richards[50] suggested that this acyl intermediate may then react with the freed amino group of H_2N-Y to form the covalent amino intermediate discussed above. Hence, this intermediate was proposed to be central to both acyl and amino hydrolysis and probably therefore to the acyl and amino transpeptidation mechanisms.

Acyl transpeptidation reaction products were observed in some cases to give much higher yields than the corresponding hydrolysis, indicating that covalent (acyl) intermediates may also be required in the hydrolysis process. The experiments of Newmark and Knowles[54] on [^{14}C]Leu-Tyr-[^{3}H]Leu showed that both amino and acyl transpeptidation occurred simultaneously with this substrate (though acyl transpeptidation dominated). They concluded that the relative importance of acyl and amino transfer probably depended only on the ease with which the amino and acyl portions leave the active site after hydrolysis. A common "intermediate" for both hydrolysis pathways is implicit in this suggestion, possibly the acyl intermediate shown above.

As for the amino intermediate, the postulated acyl intermediate in the acyl transfer could not be trapped with nucleophiles such as hydroxylamine[55] and methanol.[35a] A somewhat perplexing observation is the lack of either kind of transpeptidation with peptide substrates larger than about 7 residues. If covalent intermediates are involved in the hydrolysis of both large and small substrates, there being no reason to suppose that there is a different mechanism for the large substrates, transpeptidation would be expected. Antonov[56] suggested that the lack of (amino) transpeptidation observed with a particular substrate was due to an increase in the

rate of decomposition of the amino enzyme intermediate. It was proposed that the larger substrates give less stable covalent intermediates that are hydrolysed before they can be trapped by a suitable peptide acceptor. Rich[57] proposed that transpeptidation arose due to a slow, structure-dependent release of products from the active site. This idea was further developed by Antonov,[56] who calculated a dissociation constant for the enzyme-product complex that is consistent with the rates of transpeptidation seen (see below). Thus, the large substrates may be lost from the active site more rapidly than the smaller substrates or, again, that the covalent intermediates are less stable for large substrates.

More recently transpeptidation studies[58] have been used to investigate the incorporation of ^{18}O-label into substrate and transpeptidation products from $H_2{}^{18}O$. Extensive ^{18}O-label incorporation into early transpeptidation products challenges the suggestion[59] that the incorporation could be due to secondary reactions in the long incubation times used by Antonov.[56] Thus, these experiments on transpeptidation provide support for a non-covalent mechanism (see below). The authors also found[58] that in $H_2{}^{18}O$ only one ^{18}O atom was incorporated into the oxygens of the carboxyl group of the C-terminal cleavage product, which implies that the amide cleavage reaction is irreversible.

Other studies on transpeptidation by porcine pepsin have been undertaken by Hofmann *et al.*[60] Only acyl transpeptidation products were observed and then only after a lag of some minutes, possibly due to a requirement for the dipeptide substrates to bind productively before transpeptidation and release could occur. It was pointed out by Silver and James[61] that a lag in formation of transpeptidation products is incompatible with the formation of an acyl-enzyme intermediate that is subsequently trapped by a suitable acceptor formed during catalysis. This is because if the acyl enzyme intermediate was formed the transpeptidation products should have begun to appear immediately, with no lag.

The only piece of unequivocal evidence in favour of acyl intermediates in aspartic protease catalysis comes from the work of Kaiser[62] on sulfite ester hydrolysis. Here, the acyl intermediate could be trapped with hydroxylamine. It was shown that the same active site is used for sulfite and peptide hydrolysis, but that only one of the active site aspartates was necessary for sulfitease activity.[62b] The mechanistic

relevance of the observation of sulfitease activity to the physiological reactions of the aspartic proteases remains unclear.

1.3.2 Difficulties with the covalent mechanisms

Studies designed to trap any covalent intermediates involved in catalysis by detection of "burst" release of products at low temperature[59,63] provided no evidence for a covalent intermediate with either pepsin or penicillopepsin. The same experiments also failed to trap any intermediates by low temperature denaturation and precipitation. As has been pointed out previously[35a] any covalent intermediate must come after the rate-determining step, as all attempts to observe a "burst release" of product have failed.[64] This stands in contrast to the serine proteases where the detection of burst release of products provided good evidence for the existence of a covalent acyl intermediate.

The analysis of X-ray crystal structures of aspartic proteases complexed with transition state mimics has raised objections to the proposed covalent mechanisms. The active site cleft appears to be too small to encompass the acyl (or amino) intermediates required.[65] However, any deductions based on static crystal structures must be assessed carefully, due to the large conformational changes that the aspartic proteases are observed to undergo during catalysis[66] (see also[67]).

A major part of the covalent mechanism hypothesis hinged on the kinetics of product release. However, it has been shown by stopped-flow fluorescence that the products of hydrolysis of good substrates leave the active site at the same time.[68] Thus, with relatively poor substrates the interaction of the two products may be coupled such that the nature of one influences the rate of departure of the other, giving rise to an apparently ordered loss of products. This coupling probably occurs through effects on the conformation of the active site and does not relate in any way to the chemical mechanism.

The key experiments that have most conclusively refuted the possibility of a covalent mechanism have come from the work of Antonov[56] and Rich.[69] Using pepsin Antonov studied the incorporation of ^{18}O-label into uncleaved substrate and transpeptidation products from $H_2{}^{18}O$. The results were only consistent with a non-

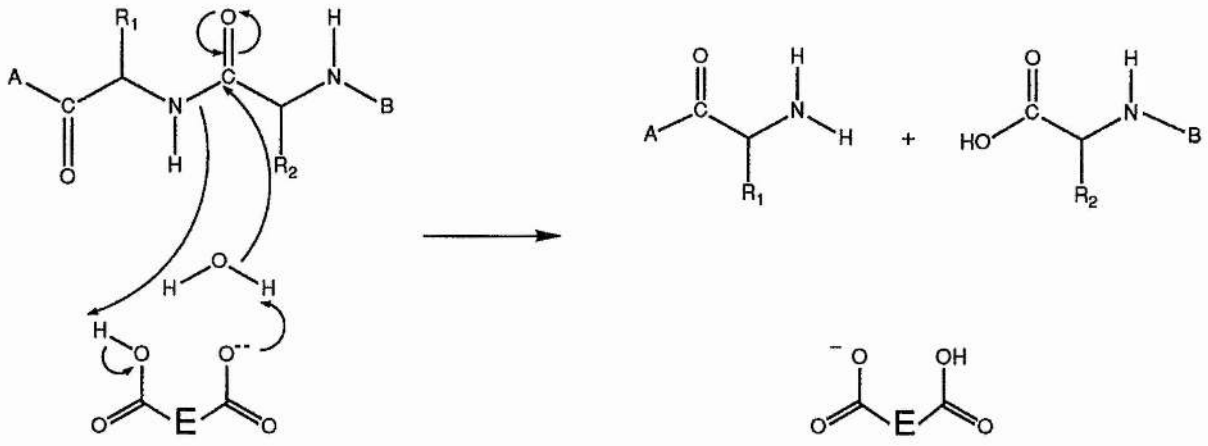
covalent, nucleophilic water mechanism. Antonov also showed that there is no incorporation of ^{18}O-label into the active site carboxyl groups of the aspartic proteases during peptide bond cleavage.[56b] Incorporation would be expected if either of the two covalent mechanisms outlined above were true, as hydrolysis of either of the covalent intermediates would lead to label becoming incorporated the active site carboxyls. Rich[69] has investigated the ^{13}C spectra of inhibitors bound to pepsin, in one case[69b] using isotope shifts induced by ^{18}O from $H_2{}^{18}O$. He has shown direct evidence for the formation of a *gem*-diol tetrahedral intermediate at the scissile carbonyl. Both sets of results are incompatible with either covalent mechanism, as hydrolysis of either of the covalent active site species mentioned above would result in ^{18}O-label incorporation into the active site carboxyls.

Another difficulty with the covalent mechanisms as presented was their inability to account of the observed electrophilicity of aspartic protease catalysis. There seemed no good rationalisation for the relative slowness of ester cleavage by the aspartic proteases compared to the serine proteases if the aspartic proteases were using a nucleophilic carboxylate in the cleavage of esters as well as amides. These mechanistic deficiencies and the consistent failure to directly confirm the existence of the covalent intermediates led to the suggestion that they may not exist.[70] However, as was pointed out by Spector in a discussion of possible covalent intermediates in aspartic protease catalysis, "the absence of evidence is not evidence of absence".[71] Attention was then turned to the development of non-covalent mechanisms, based on the concept of the nucleophilic water molecule.

1.3.3 The nucleophilic water mechanism

The nucleophilic water mechanism or general base mechanism is now generally accepted as the correct mechanism for the aspartic proteases. In this mechanism an active site water molecule is deprotonated by the active site aspartate side-chain of lowest pK_a and the nascent hydroxide ion then attacks the scissile carbonyl. The resulting tetrahedral intermediate then collapses to form the free acid and free amine hydrolysis products. The position of the water molecule in the active site and the nature of the residues involved in catalysis have been the subject of debate (see

below), but the overall features of the mechanism are outlined in Scheme 1.3.7.



Scheme 1.3.7: The nucleophilic water mechanism
where E represents the enzyme

Exchange of a proton will regenerate the correctly protonated form of the active site aspartates.

A major challenge for any non-covalent mechanism is to explain transpeptidation. The principle of microscopic reversibility shows that formation of peptide bonds at the active site of the aspartic proteases must be possible. The synthesis of peptide bonds by pepsin is central to the operation of the activator peptides (see above). At neutral pH the overall equilibrium constant for the exchange of one peptide bond for another is around 1.[56] Thermodynamically, therefore, there is no barrier to extensive transpeptidation. Kinetically the situation is less clear. It has been shown[56] that the dissociation rate constant for the enzyme-product complex must be on the order of 10 s^{-1} to allow time for an acceptor to be bound and an amide bond to be formed between them. How this huge (10^3-10^4 fold) decrease in the dissociation rate for an enzyme-substrate complex is achieved is not known.

High resolution X-ray crystallographic studies on various aspartic proteases[72] show electron density located between the two active-site carboxyl groups. Most groups have ascribed this density to a water molecule, due to its constancy under various crystallisation conditions. It was therefore proposed that this water molecule is the lytic water molecule illustrated above. An alternative proposal, by James,

postulated a different water molecule as the nucleophile (see Section 1.3.3.1).

Which of the two aspartates has the side chain of lowest pK_a has been difficult to identify (see Section 1.2.3). It was wrongly assigned initially to Asp-32 on the basis of chemical labelling experiments. These were shown later to be faulty as different conditions produced different labelling patterns. James,[73] Newman *et al.*[74] and Suguna *et al.*[72b] have found that the electron density assigned to the water molecule at the active site was located slightly closer to Asp-32 than to Asp-215, indicating that Asp-32 was deprotonated. However, the same effect is not seen in other structures[75] and the difference is only barely statistically significant in the structures in which it is found.

Theoretical approaches have been used in an effort to resolve this problem and early quantum mechanical calculations[76] were ambiguous, some indicating a small preference for ionisation of either Asp-215[76a] or Asp-32 [76b] in renin. Later studies[77] showed that the difference in energy between the two aspartates being ionised was small for the fungal proteases. Goldblum[77a] found a consistent preference for ionisation of Asp-32, in accord with the X-ray structure interpretation.[72b] However, Turi *et al.*[77b] found a preference for ionisation of Asp-215, influenced by the protonation state of the adjacent Asp-307. The proton may therefore be shared equally by the aspartates *i.e.* it shuttles quickly back and forth between the two carboxyl groups, possibly by quantum mechanical tunnelling if the two inner oxygen atoms are close enough. The inter-oxygen distance is usually about 2.9 Å in the crystal structures, but may vary greatly during the catalytic cycle. However, the path for proton transfer may not be simply along the line joining the carboxyls and may therefore have a high activation energy. Thus, some enzyme molecules may have Asp-215 ionised and the rest Asp-32 ionised, with the state changing with difficulty, if at all.

There are several mechanistic hypotheses based on the general base concept outlined above. They share some common features and differ over the same questions. The shared features are those central to the general base mechanism, the involvement of the active site aspartate dyad and a nucleophilic water molecule. The differences are found in:

i) the proposed protonation state of the tetrahedral intermediate.

ii) the timing of proton switches in the breakdown of this intermediate.

iii) the direction of binding of the scissile carbonyl with respect to the aspartate dyad.
iv) the nature of the Brønsted/Lewis acid component.

The mechanisms will be listed, then examined and critically evaluated in turn. Six distinct hypotheses have been proposed by:
a) James, based on the high-resolution structure of penicillopepsin complexed to pepstatin.[78]
b) Polgar, the so-called "push-pull" mechanism, involving simultaneous protonation/ deprotonation events in the formation and breakdown of the tetrahedral intermediate.[79]
c) Pearl,[80] where binding of side-chains on both sides of the scissile bond causes rotation of the scissile amide bond out of planarity.
d) Blundell, postulating that the active site is evolved to stabilise a negative charge on one of the aspartates and not on the tetrahedral intermediate, as the other proteases are.[21]
e) Davies,[72b,81] which is a mechanism similar to Pearl's.
f) Hemmings,[82] in which the nitrogen and not the oxygen of the scissile bond is protonated prior to nucleophilic attack by the lytic water molecule. This mechanism has not yet been fully formulated and will not be discussed further.

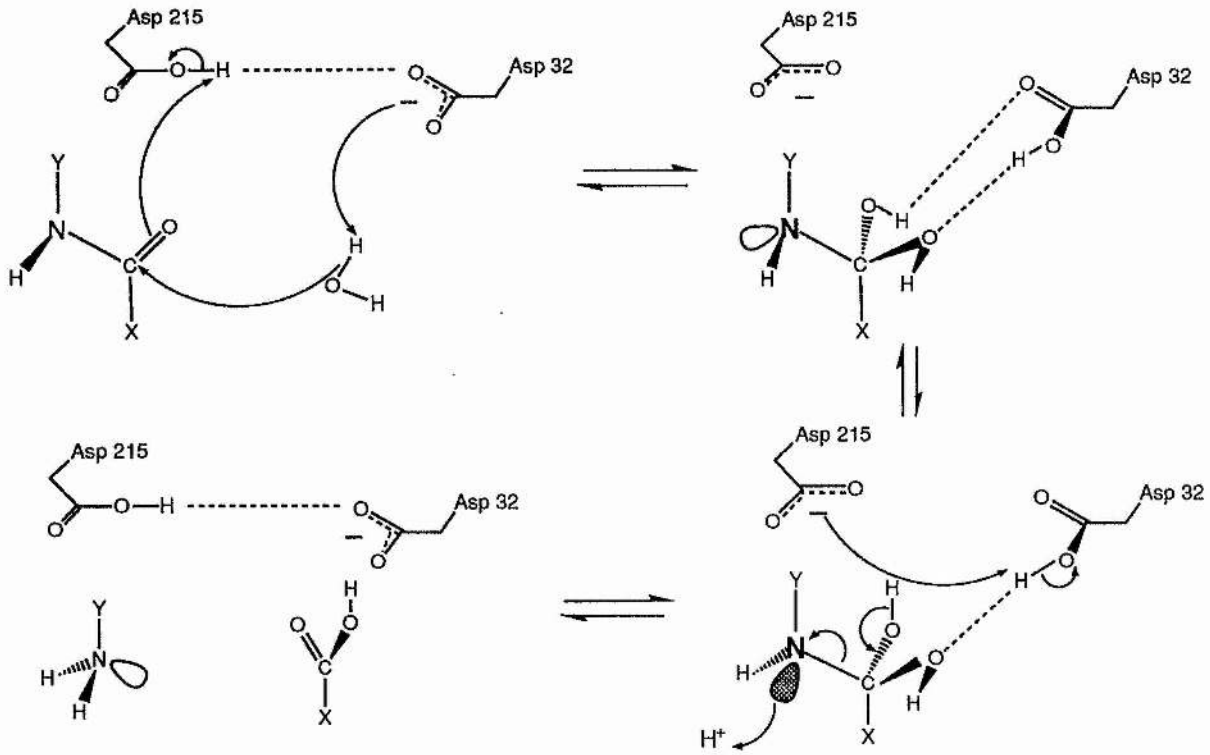
1.3.3.1 The mechanism of James

James derives his mechanistic model from the analysis of a high-resolution X-ray structure of pepstatin (Figure 1.0.1) bound to endothiapepsin. The analysis begins from a different point of departure to the other contending theories. In the course of heavy metal soaking experiments the site between the two active site carboxyl groups was found to be favourable for cation binding.[78] It was considered unlikely that this site becomes occupied, even transiently, by a moiety that becomes negatively charged during the catalytic cycle. Thus, the electron density seen between the active-site aspartate side chains is said not to be a water molecule, but rather an ammonium or hydronium ion. He also stated that this group would be displaced by the scissile carbonyl oxygen of an incoming substrate molecule. He proposed that the nucleophilic water is the crystallographic water molecule O-284,

hydrogen bonded to the side chain of Asp-32 only. It lies out of the plane of the scissile carbonyl and the two active-site aspartates (and consequently closer to the required angle of attack on the scissile carbonyl of 109°). He presents the Brønsted/Lewis acid as the proton bonded to Asp-215, the proton donor as bulk solvent (or the side-chain of D-32) and O-284 as the nucleophile. The scissile carbonyl thus points toward the active site aspartates.

The first step in the mechanism is the protonation of the scissile carbonyl by the proton shared between the two aspartates, possibly concerted with the attack on the scissile carbonyl by O-284, deprotonated by Asp-32. The tetrahedral intermediate formed is neutral. Protonation of the amine leaving group may occur by either of two routes. An inversion of configuration at the nitrogen will allow reprotonation from the bulk solvent, as shown below. Alternatively, a proton relay *via* Asp-215 from the innermost hydroxyl group of the *gem*-diol intermediate could be used. Nitrogen inversion is a very rapid process[83,84] so should have no effect on the rate of catalysis.

An important objection to the mechanism as proposed is that although the site between the active site aspartate residues is a cation binding site *in the native enzyme* this is not to say that this is a cation binding site during the catalytic cycle. Recent quantum mechanical calculations[77b] have found that the more stable active site arrangement is for this moiety to be a hydronium ion, not an ammonium ion. However, the same study found that the most energetically favoured form for the active site is to bind a water molecule, not a hydronium ion.



Scheme 1.3.8: The mechanism of James

The requirement for prior protonation of the scissile carbonyl is shared by acid-catalysed water hydrolysis of amides (see Section 1.5). The pK_a of an amide oxygen is around 0 to -2[85] so assistance by water attack may be required.

Another fundamental difficulty with this mechanism is that it is based upon a crystal structure in which the inhibitor is not congruent with a substrate. The pepstatin inhibitor binds in a manner that is most unlike that of a substrate, especially in the crucial P_1-P_1' region. Its P_2' side-chain binds not in the S_2' subsite, but the S_1' subsite, due to a curl in the inhibitor backbone. More substrate-like inhibitors do not show this deviation. Thus, detailed deductions about substrate binding and conformation based on this structure are difficult to make confidently. Other crystal structures with inhibitors that are more congruent with peptide substrates also show that the proposed nucleophilic water, O284, is displaced by the P_1 side chain.[72b,81] Very recently this mechanism has been retracted in the light of close examination of crystal structures of penicillopepsin binding difluorostatine-based inhibitors.[87] In this paper James proposes a mechanism essentially identical to that of Davies (see 1.3.3.5).

In all the mechanisms discussed below the nucleophile is considered to be the

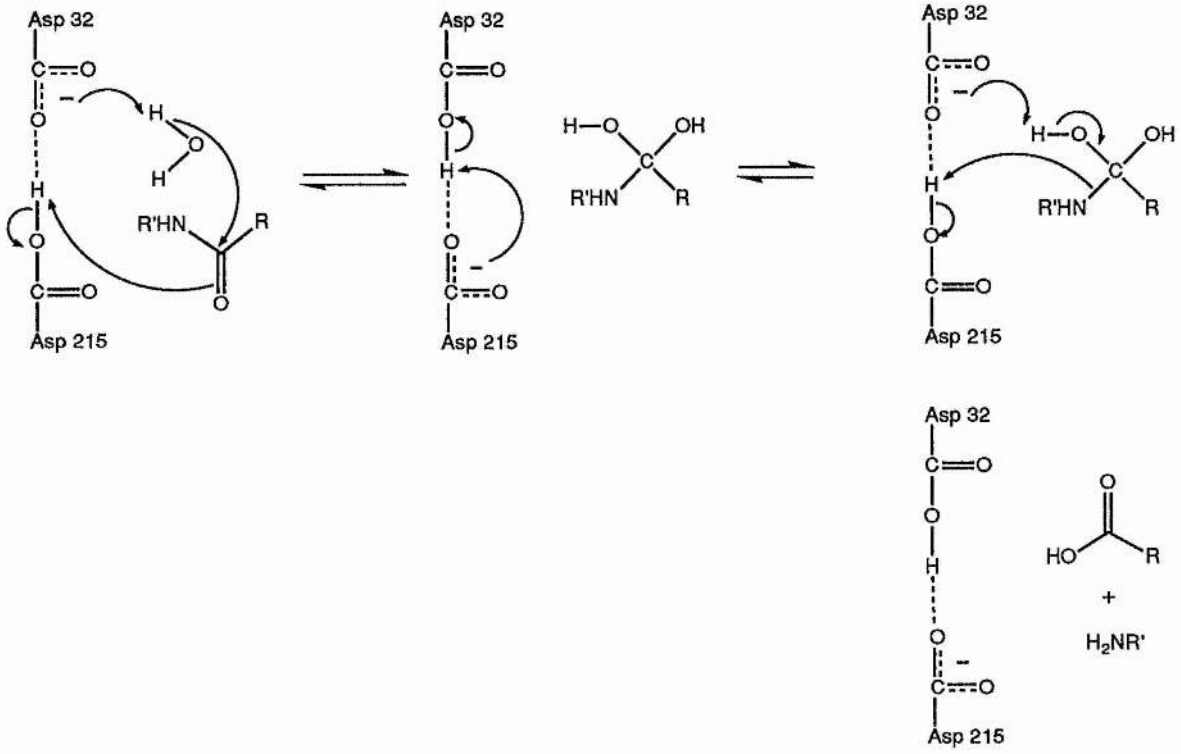
water molecule hydrogen-bonded between the active site aspartates.

1.3.3.2 The mechanism of Polgar

Polgar's push-pull mechanism involves a series of simultaneous proton transfers, as shown in Scheme 1.3.9, with the active site dyad of aspartates acting as a functional unit. He presents the Brønsted/Lewis acid as the proton bound between the active site aspartates and the proton donor as D-215. This requires the scissile carbonyl to point toward the active site aspartate residues.

The mechanism proceeds as follows. First, deprotonation of the nucleophilic water molecule by D-32 and attack on the scissile carbonyl is paralelled by protonation of the scissile carbonyl oxygen to give a neutral *gem*-diol intermediate. Subsequent deprotonation of this tetrahedral intermediate and reformation of the carbonyl group is synchronous with C-N bond cleavage and protonation of the leaving group nitrogen. The requirement for the scissile carbonyl to point toward the aspartate dyad means that, if the scissile amide is planar, the nucleophilic water molecule will be displaced. However, if the scissile amide is distorted from planarity then there is available space for the water molecule to remain bound between the aspartate side-chains. Hence, it bears a close resemblance to the mechanism advocated by Pearl in Scheme 1.3.3.3 below.

Both parts of this mechanism are symmetrical, as both are facilitated by push-pull general acid catalysis.



Scheme 1.3.9: The mechanism of Polgar

1.3.3.3 The mechanism of Pearl

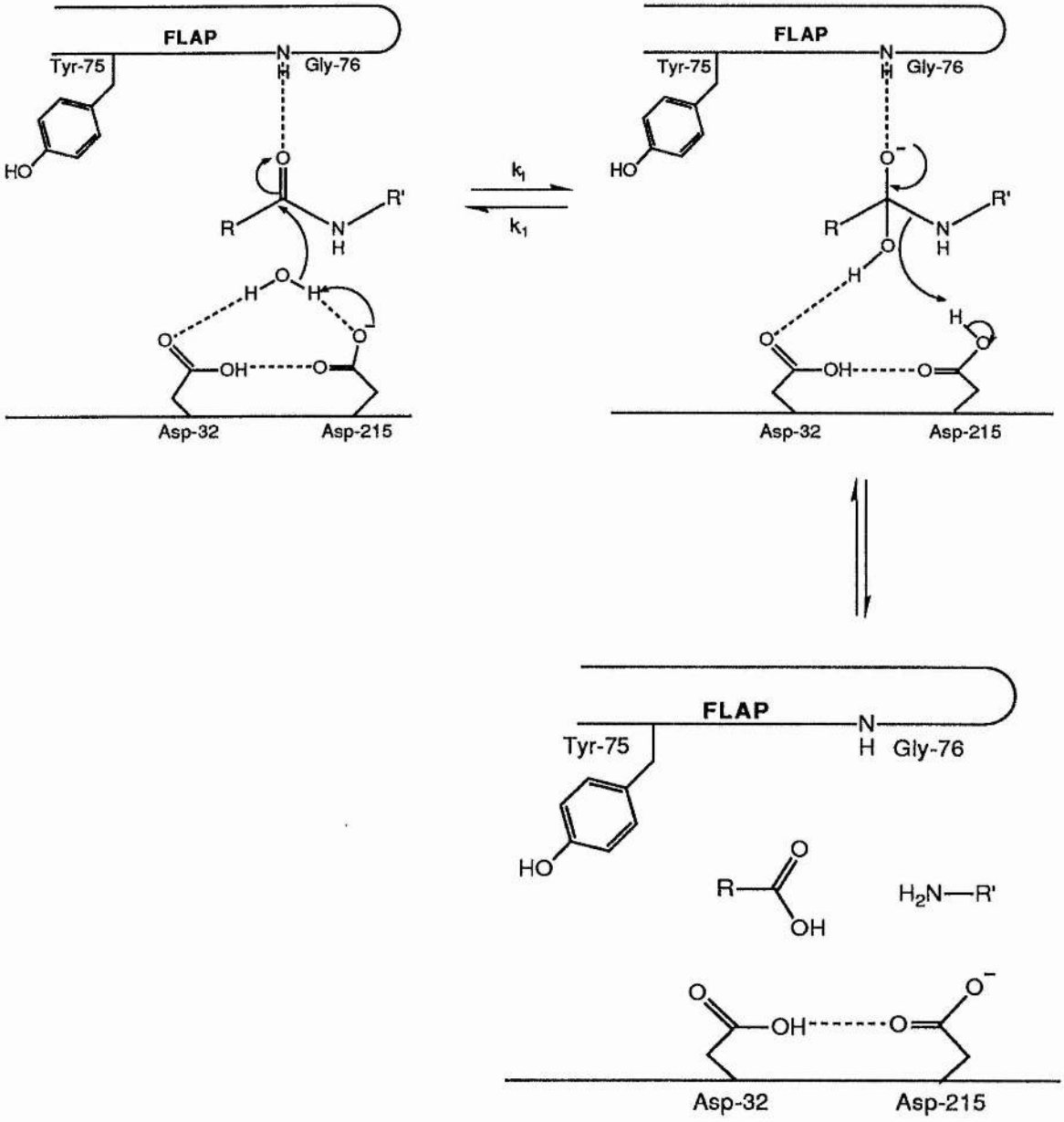
The key part of the mechanism of Pearl is the distortion of the scissile peptide bond on binding, leading to a reduction of scissile amide bond resonance. He presents the Brønsted/Lewis acid component of catalysis as hydrogen bonds from residues in the flap (Gly-76 and possibly Tyr-75 in pepsin) to the scissile carbonyl and the proton donor as Asp-215. In this mechanism the scissile carbonyl points away from the active site aspartates and the tetrahedral intermediate carries a negative charge, which is stabilised by hydrogen bonds from the flap.

Reduction of amide resonance increases the electrophilicity of the scissile carbonyl and the basicity of the amide nitrogen. Increasing the carbonyl electrophilicity increases the rate of tetrahedral intermediate formation and increasing the basicity of the amide nitrogen increases the rate of breakdown to products. As mentioned above an important feature of aspartic protease-catalysed peptide hydrolysis is the marked increase of k_{cat} with the length of the peptide, with little or no change in K_M.[29a] This mechanism was developed to account for these observations.

The extra binding energy available from the secondary binding of substrate side-chains is utilised to lower the activation energy of the reaction by distorting the scissile bond from planarity. The enzyme therefore destabilises the ground state of the reaction. It is likely that strain is transmitted through the enzyme rather than through the substrate, that is the binding pockets containing the side-chains move relative to each other. A similar proposal was first made by Fruton some 20 years earlier.[87]

This model has received a measure of experimental support, with several inhibitors having been found to have their scissile bond mimics distorted from planarity. Reduced peptide based inhibitors of rhizopuspepsin[72a,b] and the HIV-1 protease[88] show the scissile bond mimic to be significantly twisted from planarity. Distortion has also been observed in residues flanking the scissile bond[89] in other classes of inhibitors. It is known that distortion of an amide group from planarity greatly increases the susceptibility of the carbonyl to nucleophilic attack by hydroxide ions,[90] as well as increasing the basicity of the nitrogen.[91]

Pearl also suggests that the oxyanionic tetrahedral intermediate is stabilised by an interaction with the edge of the aromatic ring of Tyr-75, which is conserved in all mammalian and fungal proteases so far sequenced.[21] Thus, this mechanism also postulates some similarities between the serine and the aspartic proteases. This interaction between an oxyanion and the aromatic ring was suggested by Blundell as a possible mechanism for stabilisation of the oxyanion intermediate.[92] It has been estimated[93] that this kind of interaction can stabilise the oxyanion by as much as 5 kcal mol^{-1}. However, while Tyr-75 is invariant in monomeric aspartic proteases it is absent in retroviral proteases,[94] so it is unlikely to be a pivotal feature of the mechanism of all aspartic proteases. Also, Suguna *et al.*[72b] could not model the proposed hydrogen bond between NH of Gly-76 in the flap and the scissile carbonyl.

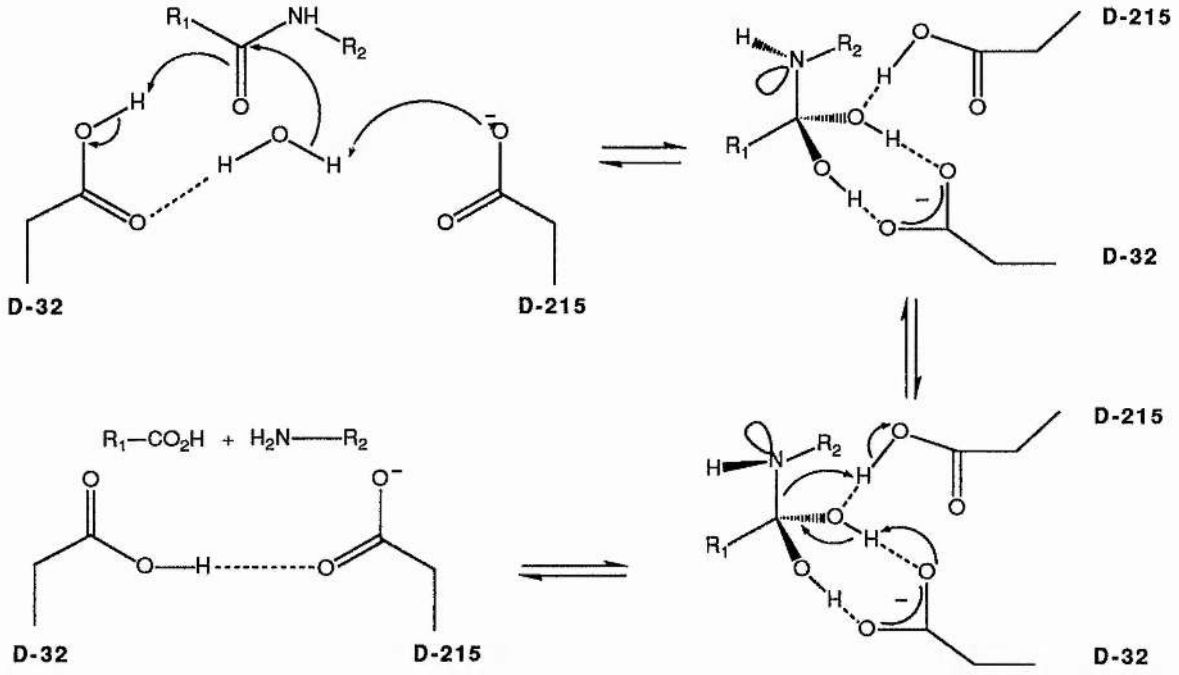


Scheme 1.3.10: The mechanism of Pearl

1.3.3.4 The mechanism of Blundell

The mechanisms of Blundell and Polgar share the important point that the tetrahedral intermediate is neutral *i.e.* protonated, and a function of the aspartate pair is to facilitate this protonation and subsequent deprotonation. However, Blundell sees little similarity between the mechanisms of the serine and aspartic proteases,

particularly with respect to the existence of an oxyanion hole. This is of central importance in the serine, cysteine and metalloproteases and a number of mechanisms have implicated a similar structure in the aspartic proteases.



Scheme 1.3.11: The mechanism of Blundell

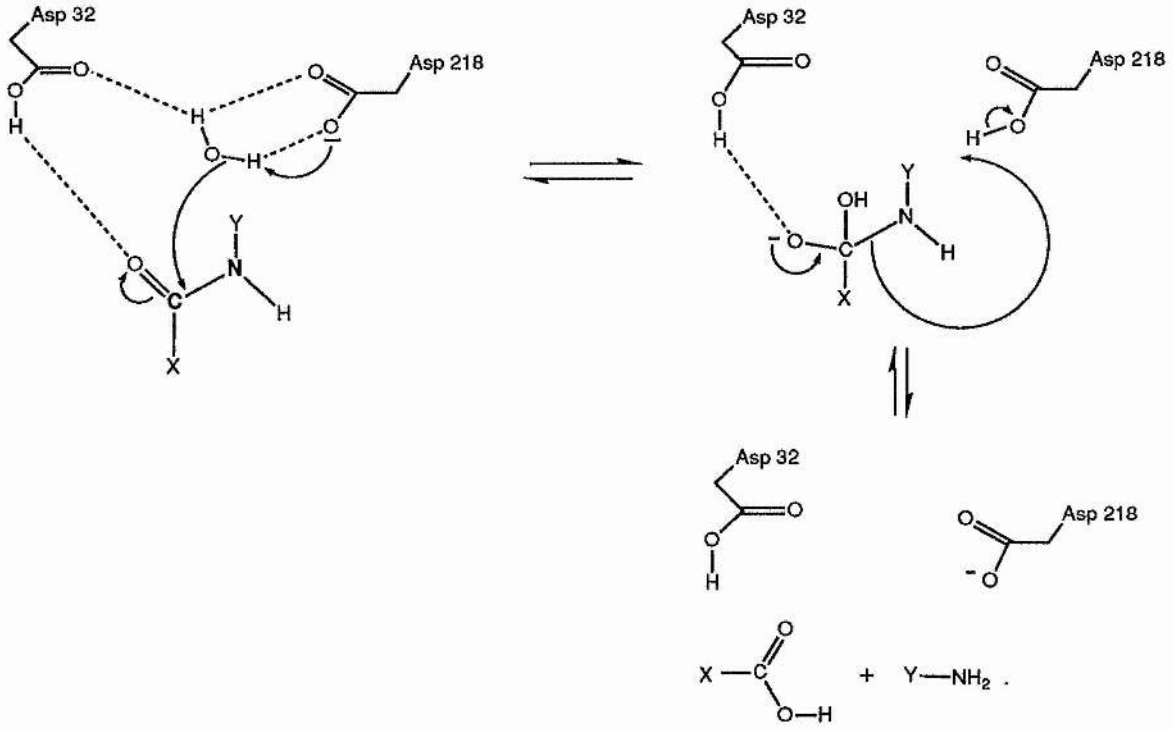
The key point is the inversion at nitrogen followed by rotation about the $C(OH)_2$-N bond to allow protonation from D-215. This serves the dual purpose of destabilising the tetrahedral intermediate by removing a hydrogen bond and activating the amine leaving group by protonation. The ionisation states of the aspartates shown are supported by a quantum mechanical study on a model system[95] showing that deprotonated D-32 is around 5 kcal mol^{-1} more stable than the alternative, deprotonated D-215. There is no direct electrophilic activation in the mechanism, in contrast to the mechanisms of Pearl and James. The lack of electrophilic activation by protonation (as in the mechanism of James) is supported by quantum mechanical studies[96] showing that the distance between the scissile carbonyl and the proton of D-215 is too large for significant interaction. However, the same study indicated that there is no possibility of nitrogen being protonated from D-215 during the breakdown of the tetrahedral intermediate as they are too far apart. Rotation about the bond

bearing the carboxyl group of D-215 is required first.

The hypothesis states that the aspartic proteases stabilise the negative charge formed during the cleavage reaction by localising it on the aspartate pair and not on the tetrahedral intermediate as the serine, cysteine and metalloproteases do. This was considered to be supported by the observation of frequent replacement of S-218 (in pepsin numbering) by Ala in the aspartic proteases of a higher pH optimum, *e.g.* renin and HIV-1 PR. This replacement means that the anion on D-32 is less stabilised by hydrogen bonding and so the optimum pH of an aspartic protease with Ala at position 218 will be higher as the anion is more difficult to stabilise at lower pH. This assumption was supported by semi-empirical calculations from the group of Goldblum[97] on the HIV-1 protease active site. However, recent work by Lin *et al.*[26,98] has shown that this is not so (see Section 1.2.3), replacement of the serine/threonine residue after the D-T-G active site triplet in pepsin and rhizopuspepsin by alanine has no effect on the pH-rate profile of either enzyme. Thus, this mechanism too has little experimental support.

1.3.3.5 The mechanism of Davies

The mechanism of Davies bears a close resemblance to that of Pearl. The intermediate is postulated to be an oxyanion, stabilised by a hydrogen bond from the OH of D-32. However, the electrophilic component here is provided by hydrogen bonds to the scissile carbonyl oxygen from OH of D-32 and OH of S-38 and not by residues in the flap. The proton donor is D-218, the lone pair on the nitrogen is localised, pointing toward it, by a weak hydrogen bond from the carbonyl of the P_2 residue to the NH of the scissile bond (not shown below).



Scheme 1.3.12: The mechanism of Davies

Which of these mechanisms most truly reflects the chemistry occurring at the active site of the aspartic proteases has important ramifications for the design of inhibitors. If the tetrahedral intermediate is negatively charged then anionic inhibitors should be tightly bound. However, if the intermediate is neutral then anionic species will be poorly bound and uncharged inhibitors will function best. This point is discussed further in Section 1.6.3.

1.3.4 Importance of the S_3 site

Holladay *et al.*[57] reported that the occupation of the S_3 subsite in pepsin is necessary for the release of transpeptidation products, thus transpeptidation may only occur with (small) substrates that do not occupy the S_3 subsite. The observed delay in release of transpeptidation products noted above may thus be caused by the time needed to build up the substrate by transpeptidation to such a size that it can occupy the S_3 subsite. The importance of interaction of the substrate with the S_3 site is underlined by the approximately 500 fold increase in k_{cat} for substrates having

a P_3 residue compared with otherwise identical substrates lacking the P_3 residue (at constant K_M).[99] This could be due to distortion of the substrate, non-productive binding of a shorter substrate or a conformational change in the enzyme that results in more efficient catalysis. Non-productive binding does not affect k_{cat}/K_M when Michaelis-Menten kinetics are obeyed, as they are for the aspartic proteases, so this is not a valid explanation. The idea that a conformational change is involved is supported by the observation of multistep binding for inhibitors with a P_3 residue and single step binding for those without.[100] The same effect is seen in the difference in K_i values between the pepstatin analogues shown below, where the statine residue (Sta) occupies the S_1' pocket.

Table 1.3: *Inhibition constants for statine based inhibitors of varying size*

I	Iva-Val-Sta-OEt	$K_i > 10^{-4}$ M
II	Iva-Val-Val-Sta-OEt	$K_i = 2.4 \times 10^{-8}$ M
III	Iva-Val-Val-Sta-Ala-Iaa	$K_i = 6.5 \times 10^{-9}$ M
IV	Iva-Val-Sta-Ala-Iaa	$K_i = 7.6 \times 10^{-6}$ M

Analogue II has a isovaleryl (Iva) residue at the P_3 position and this difference gives rise to the four order of magnitude increase in inhibition over analogue I. Similarly for III and IV, occupation of the S_3 pocket increases the inhibition of penicillopepsin around 3 orders of magnitude.[101]

The most important residue for substrate binding in the S_3 pocket is Thr 218 (in pepsin numbering), which is conserved as Ser or Thr in every monomeric aspartic protease so far sequenced.[21] It accepts a hydrogen bond from the amide NH of the P_3 residue and donates a hydrogen bond to the CO of the P_3 NH,[86] so these hydrogen bonds are likely to be constant for all substrates. In renin these interactions are absent,[102] so the occupation of the S_3 pocket will be less important. The amount of hydrophobic interaction the P_3 side-chain makes with the S_3 pocket is variable,[103,104] so the effects of binding a side-chain in the S_3 pocket must arise

from the hydrogen bonds formed. It may be that formation of these hydrogen bonds is important in triggering a conformational change that facilitates cleavage or loss of product.[30] This conformational change may be the rigid body movement observed in some crystal structures of aspartic proteases complexed with inhibitors (see Section 1.4.1).

The importance of the S_3 site is further demonstrated by the observation that certain substrates will change their binding mode and shift the scissile bond if that places a residue in the S_3 pocket.[30] Thus, Ac-Ala-Lys-Phe(NO_2)-$(Ala)_2$-NH_2 incubated with penicillopepsin will be cleaved between the Phe(NO_2) (nitrophenylalanine) and Ala residues and not the favoured cleavage site between Lys and Phe(NO_2) as this ensures that the S_3 pocket is filled by the Ala side-chain. This drive to occupy the S_3 site may explain the phenomenon of the activator peptide, a small non-substrate peptide that, when added to an incubation, increases the rate at which a substrate peptide is cleaved.[105] The peptide may either form a transient condensation product with the substrate or may simply occupy the S_3 site while the substrate occupies S_1 and S_1'. In either case a species that occupies the S_3 site is central to the action of the activator peptide. Silver and James[105a] showed that the key step in activator peptide function is pepsin catalysed condensation of the activator and substrate peptides. The resulting large peptide is then cleaved at a different bond, much more rapidly than the substrate peptide. This implies an as yet unexplained difference in specificity for synthesis and cleavage of peptide bonds by monomeric aspartic proteases.

1.3.5 Solvent isotope effects

The nucleophilic water proposal leads obviously to the use of solvent deuterium isotope effects to probe the mechanism. Early experiments using this technique gave very varied results and were of little use in untangling the important steps in catalysis. The early work found no isotope effect[106] using small peptides and phenyl methyl sulfite. However, an effect was found using $(Gly)_3$-Phe(NO_2)-Phe-OMe.[107] Hunkapiller and Richards[50] also found a large (3.0) solvent isotope effect (on V_{max} and not K_M) in the hydrolysis of N-trifluoroacetyl-L-phenylalanine by pepsin, the largest so far reported.

Recently, Cunningham *et al.*[108] found a different rate-determining step in the cleavage of large and small substrates, leading to solvent isotope effects that vary with the size of the substrate. They observed no isotope effect on the cleavage of Ac-Lys-Phe(NO_2)-NH_2 and an isotope effect on V_{max} of 2.1 on the cleavage of Ac-Ala-Ala-Lys-Phe(NO_2)-Ala-Ala-NH_2. Proton inventory studies showed that two or more protons were involved in the rate-determining step for the large substrate. This step was proposed to be the formation of two hydrogen bonds at the S_3 subsite (and possibly at the S_2' site also). Significant conformational changes were observed, by fluorescence, during the cleavage of the large but not the small substrate, which they suggested was the rigid body movement observed by Sali.[100] This movement may be triggered by the occupation of the S_3 site and the formation of the hydrogen bonds between substrate P_3 NH and O^{γ} T-218 and P_3 CO and NH T-218. However, this is not consistent with the solvent isotope effect observed for the small substrate N-trifluoroacetyl-L-phenylalanine, which cannot occupy the S_3 site. The rate determining step for the small substrates was postulated to be the loss of products.

More recent experiments on renin,[109] pepsin[110] and chymosin[111] have shown a solvent isotope effect on V_{max} but not on V_{max}/K_M with a series of hexapeptide or larger substrates. V_{max} isotope effects involve all steps after substrate binding, while V_{max}/K_M effects relate to all steps before the first irreversible step. As the effect is on V_{max} only and transpeptidation shows that catalysis is readily reversible (see Section 1.3.1), the isotopically-sensitive step is likely to be the release of the first product. This is in accord with the (partially) rate-determining step being product release, as was proposed from transpeptidation experiments. This analysis is at variance with the picture of aspartic protease catalysis put forward by Fruton,[29a] where a rapid equilibrium model is proposed *i.e.* all the steps along the catalytic cycle are in equilibrium with each other.

1.3.6 Summary

The mass of data accumulated on various aspects of aspartic protease structure and mechanism has not been matched by a development of theories to underpin the data. A chemical mechanism very similar to that of Blundell[21] describes well the structural side of the aspartic protease mechanism and has some theoretical support.

However, this mechanism does not explain several aspects of aspartic protease catalysis, most notably the origin of the solvent isotope effects and transpeptidation, as it has no kinetic component. A major difficulty in this area is the lack of good experimental support for any of the proposed mechanisms.

The mechanism of Blundell suggests that the tetrahedral intermediate in hydrolysis is neutral and this is borne out by quantum mechanical calculations.[31] This neutral intermediate is also the central intermediate in acid-catalysed amide hydrolysis (see Section 1.5). However, in contrast to the enzymic reactions shown above, the acid-catalysed mechanism requires the departing nitrogen to be fully protonated before the tetrahedral intermediate begins to break down.

1.4 Structure of the aspartic proteases

The first crystalline protein analysed by X-ray crystallography was pepsin[112] and since then X-ray crystal structures have been reported for a large number of the aspartic proteases. Penicillopepsin,[113] rhizopuspepsin,[72a] endothiapepsin,[100,114] pepsin,[115] pepsinogen,[116] chymosin,[117] mucorpepsin[118] and renin[119] have all had their structures solved at high resolution. These crystal structures have revealed many interesting features in the global structure and the active sites of the aspartic proteases.

It is conventional to describe the aspartic proteases as being folded into two structurally similar lobes, an N-terminal lobe and a C-terminal lobe. The lobes are related by a pseudo-dyad axis running between the two active site aspartate residues and between the two central strands of a large six strand ß-sheet that constitutes the core of the enzyme. However, since the interface between the lobes consists of a six strand ß-sheet, the structures could well be described as being made up of two mainly ß-sheet domains packed onto the interlobe ß-sheet with a small α-helix at each end. There are also some other small α-helical portions, but these are not well conserved between the different proteases. The α-helical segments are in variable positions within the structures and move to accommodate changes in crystal packing. They do not, therefore, seem integral to the structures.[120]

The N-terminal lobe (residues 1-171) is the larger of the two domains and has a

more extended structure than the C-terminal lobe (172-326 in pepsin). Within the lobes there are 2 regions forming compact domains possessing an intra-domain axis of symmetry, the C-terminal domain (S 185 to R 307, 123 residues) and the N-terminal domain (E 7 to Q 148, 142 residues) for rhizopuspepsin. The intra-domain dyad axes are approximately coplanar with the inter-lobe axis and are roughly related by it. The domains may be the remnant of the original structure of an ancestral protease that underwent two gene duplications to give the four part structures seen today.[121] There is a long (25-30 Å) cleft between the lobes that contains the active site. The two catalytic aspartates are located at the end of two psi-type loops that extend from each domain into the cleft and are related by the pseudo-dyad axis.[122] The two aspartate residues are coplanar and rigidly held by an extensive network of hydrogen bonds, both inter and intra domain, including the unique 'fireman's grip'.[122] This refers to the hydrogen bonding between the two conserved hydroxy-amino acids in the active site sequence Asp-Thr-Gly-Ser/Thr, each hydroxyl group hydrogen bonding to the other's amide nitrogen (see Figure 1.4.1). The resulting rigidity is reflected in the low isotropic temperature factor (B factor) for the active site residues of chymosin, averaging 12Å^2, compared to the average of 24 Å^2 for the whole molecule.[119]

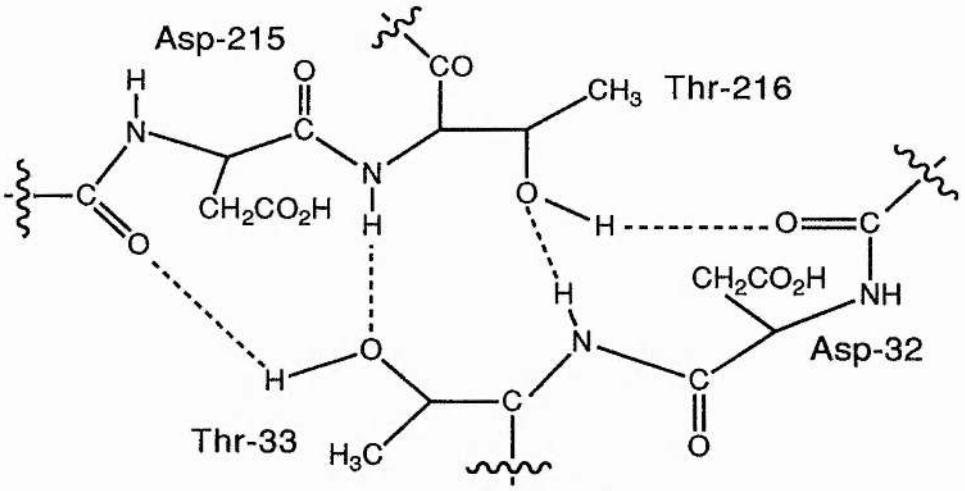


Figure 1.4.1: The "fireman's grip" in pepsin[122]

As shown below (Figure 1.4.2) the remainder of the active site is also extensively hydrogen-bonded, resulting in the entire active site being extremely rigid. This combination of conserved sequence and high levels of hydrogen bonding means

that the active sites of many of the aspartic proteases have very similar three-dimensional structures. Sielecki *et al.*[123] found an r.m.s. difference of only 0.24 Å between 105 main chain atoms from the active site of three fungal proteases and a 0.45 Å difference between these three enzymes and human renin.

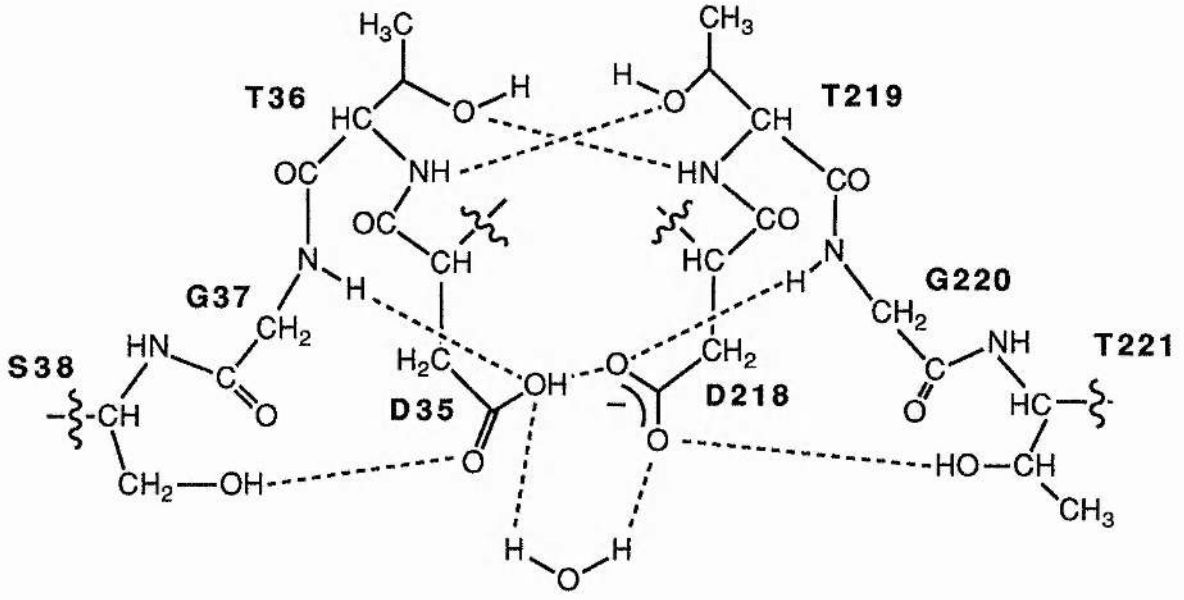


Figure 1.4.2: The hydrogen bonding network at the active site of rhizopuspepsin[72b]

The overall fold of the aspartic proteases is shown in Figure 1.4.3 overleaf.

1.4.1 Structures of aspartic proteases with bound inhibitors

These structures have shed much light on the binding mode of substrates and their interactions with the active site and the changes induced in both enzyme and substrate upon binding. The nature of the changes in the structures of the aspartic proteases on binding of inhibitors (and by extension substrates) and the relation of these changes to catalysis was for some time not well understood. The first high-

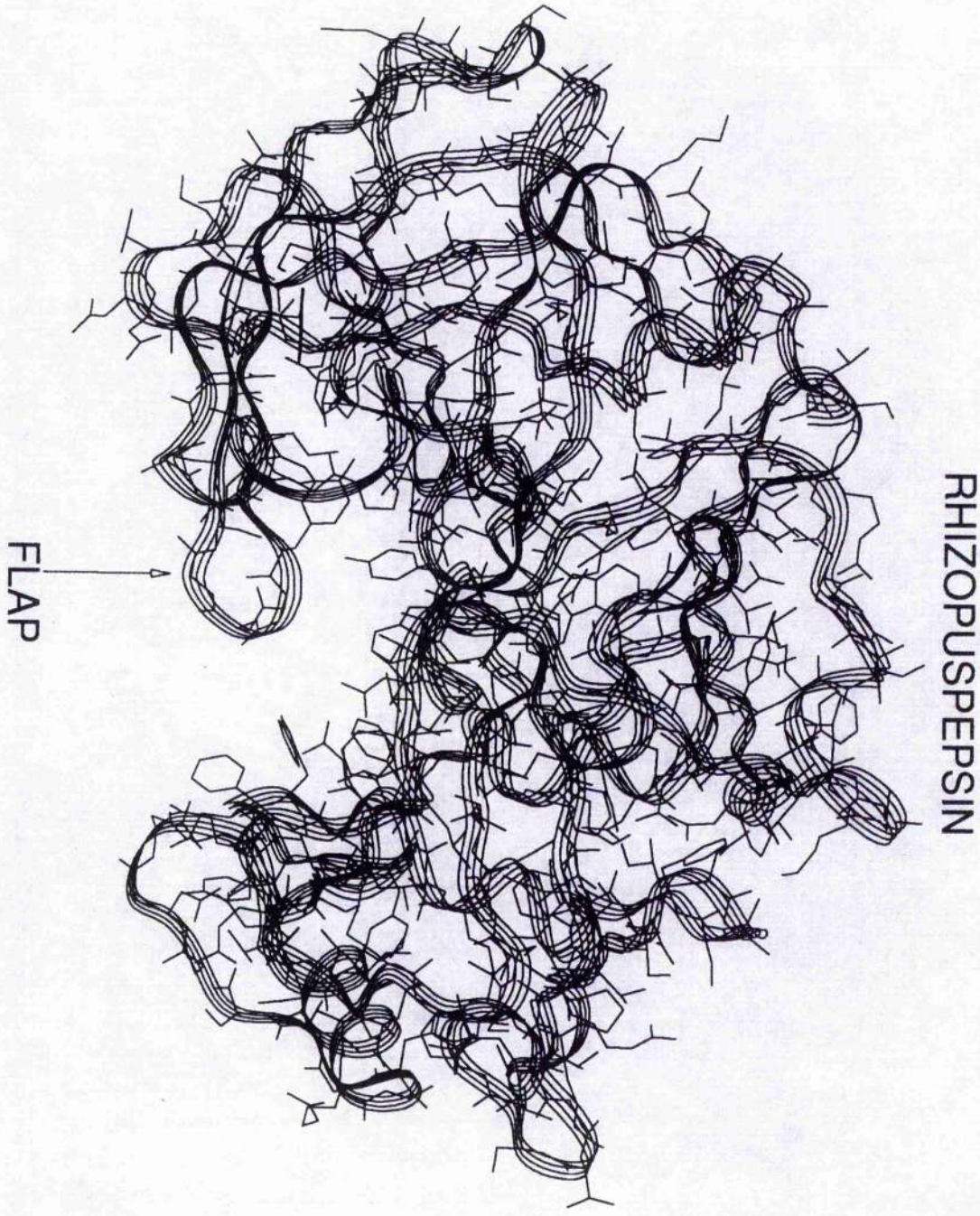


Figure 1.4.3: The overall fold of the monomeric aspartic proteases

resolution structures with bound inhibitors showed several interesting features of the binding. There is a stretch of antiparallel β-sheet between residues P_3 and P_1 of the inhibitor and residues 217-219 of the C-terminal lobe of the enzyme. There is no equivalent structure in the N-terminal lobe, with the only hydrogen bond formed between P_2' NH and the carbonyl of residue 34. Also observed was the large movement of a region known as the flap.[124] This ß-hairpin bend structure (residues 72-81 in rhizopuspepsin) folds over the active site, shielding the active site from solvent and providing further binding interactions to the inhibitor. The P_1 and P_1' residues of the inhibitor are totally shielded from solvent as a result of this movement.[89] The main hydrogen bonding interactions between a peptidic inhibitor and the monomeric aspartic proteases are shown in Figure 1.4.4 below.

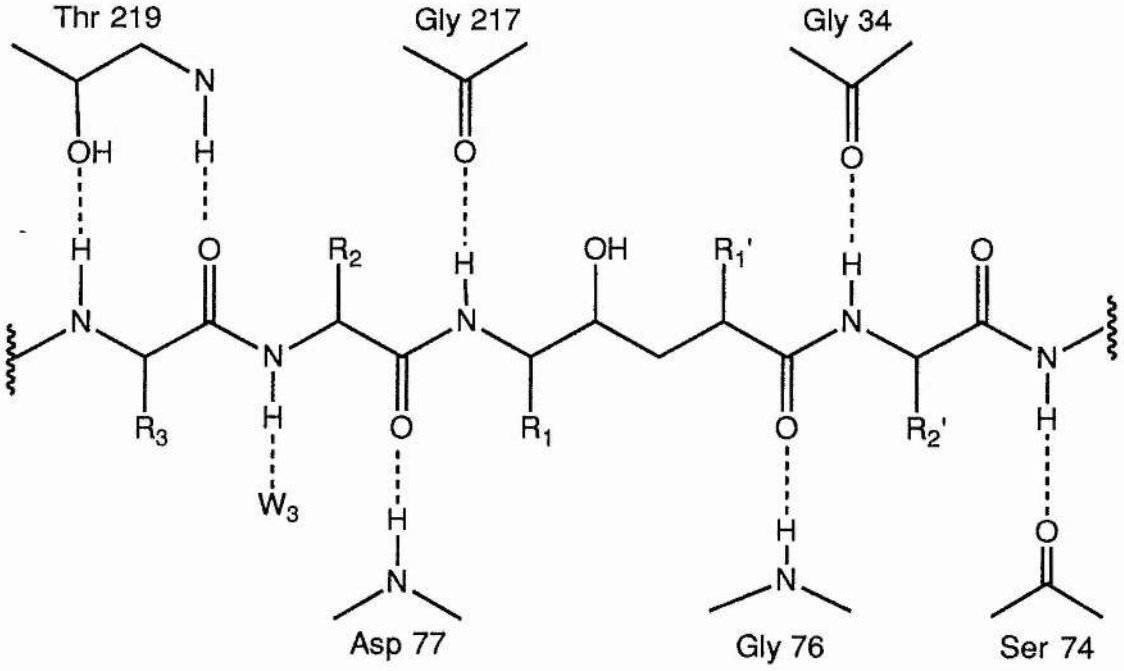


Figure 1.4.4: Hydrogen bonding pattern between the main-chain of an inhibitor and endothiapepsin[124]

where W_3 is a water molecule.

The extended conformation of the inhibitor seen in the X-ray structures above is reflected in inhibitor conformations observed in n.m.r. studies on inhibitors bound to pepsin.[125] The importance of the flap can clearly be seen from the above diagram as residues D-77, G-76 and S-74 are all from the flap and provide three strong

hydrogen bonds to the inhibitor main chain. The formation of these hydrogen bonds provides part of the driving force for flap closure. The amount of movement of the flap observed on inhibitor binding varies greatly between structures. The differences arise from the different initial flap positions in the unliganded proteases, caused by different packing interactions between enzyme molecules in the crystal. The flap is observed to be highly mobile in the unliganded structures (B value around 30 $Å^2$), becoming much more ordered on closing over the active site (B value about 8 $Å^2$).[126] The wide-ranging nature of the conformational changes is also seen in circular dichroism studies of pepsin interacting with pepstatin.[24]

There are other structural changes observed in some structures, apart from the closure of the flap. There have been reports of independent, rigid body motion of subdomains within the structure that are triggered by inhibitor binding. The movement of residues 190-303 of endothiapepsin when complexed with a reduced peptide inhibitor observed by Sali *et. al.*[89] was the first reported example. The same domain in pepsin was also reported to move as one rigid body, but in a very different fashion, by Abad-Zapatero *et al.*[122] Later studies by the same group on glycol-based inhibitors of pepsin[127] show a similar domain movement to that found by Sali *et al.*[89] This domain can be easily seen in the comparison of X-ray structures of unliganded with liganded proteases[100] where it is the point of most divergence between the different structures.

This flexible domain aligns well between various structures only if it is allowed to move independently of the rest of the protein. This division of the structure accords with the tripartite division of the aspartic protease structure discussed above, as the C-terminal residues 190-303 correspond to one rigid body and the other rigid body corresponds to the central motif and the N-terminal domain. The triggering of a rigid body movement was referred to above (Section 1.3.5) as the possible rate-limiting step in catalysis. This rigid body movement may provide the means for scissile bond distortion postulated by many of the mechanisms discussed above. The residues of the flexible domain are used to bind side-chains P_1', P_2 and P_4, and the other domain binds the other side-chains. Therefore, movement of the N-terminal domain relative to the C-terminal domain, will distort one set of side-chains relative to the other. This may distort the scissile bond from planarity as the P_1 and P_1' residues

flanking the scissile bond are bound by the different domains. Curiously, a corresponding change in enzyme fluorescence is not observed,[128] as would be expected for such a structural readjustment. A rigid body movement has also been postulated in the mechanism of the serine proteases.[129]

There seems to be little effect of inhibitor binding on the structure of the active site region, the aspartate residues remaining almost completely fixed. This is consistent with the rigid active site, reflected in the low B values and extensive hydrogen bonding, discussed above. The only movement seen is some disruption of the coplanarity of the aspartates.[89]

1.5 Model systems

The daunting complexity of the reactions catalysed by the aspartic proteases has led some investigators to the use of small molecule model systems to try and illuminate the nature of the chemical steps in the catalysis by aspartic proteases. The earliest efforts, by Bender *et al.*,[130] focussed on the intramolecular hydrolysis of phthalamic acid and this data was later used as the basis for a mechanistic proposal for the aspartic proteases.[131] A similar system, N-*iso*-propyl di-*iso*propylmaleamide, was later explored by Kirby *et al.*[132]

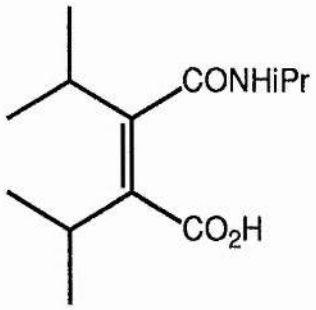


Figure 1.5.1: N-isoPropyl di-isopropylmaleamide

For a particular carboxyl substituted maleamide the pH-rate profile and the rate constant for hydrolysis were found to be comparable to those for pepsin acting on small peptide substrates. However, only when the carboxy group and the amide were in a specific structural relationship did this effect manifest itself. The relevance of this observation to the mechanism of pepsin and the other aspartic proteases was

not further explored.

The low efficiency of intramolecular general base catalysis was demonstrated in further work by Kirby[133] on the hydrolysis of monoaryl malonates and aryl hydrogen cyclopropane 1,1 dicarboxylates. Intramolecular general base catalysis is also much less efficient in water than nucleophilic catalysis.[134] These studies would seem to indicate that the aspartic proteases should utilise a nucleophilic type mechanism in order to attain the observed rate enhancements for peptide hydrolysis. However, intramolecular general base catalysis is expected to be proportionately much more efficient in low polarity environments, such as an enzyme active site.[134] The work of Gandour[135] has shown that the reactions of the model compounds used in chemical studies all involve the use of *anti* lone pairs which are much less basic than the *syn* lone pairs. Examination of crystal structures indicates that enzymic catalysis uses *syn* lone pairs.[86] The *anti* lone pairs of the catalytic aspartates are involved in hydrogen bonds, leaving the *syn* lone pairs for involvement in proton transfers during catalysis.

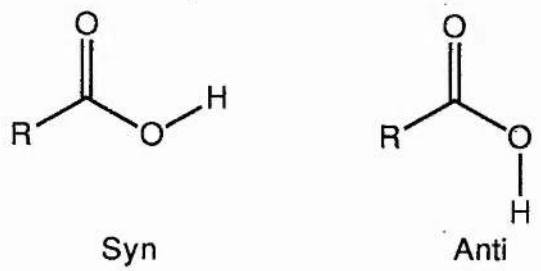


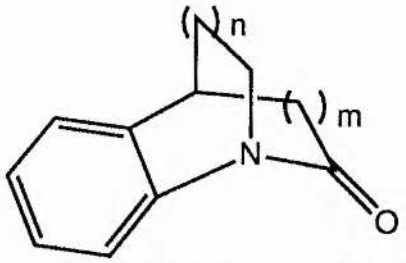
Figure 1.5.2: Syn and anti carboxyl groups

Kluger and Chin[136] studied the effect of pH and leaving group on the intramolecular hydrolysis of a variety of maleamic acids and found that the nature of the rate determining step was strongly dependent on the basicity of the departing amine. A mechanism was proposed based on the observation that the rate-determining step for hydrolysis of amides of highly basic amines is diffusion apart of the amine and the maleic anhydride. It accounts for both acyl and amino transpeptidation by postulating a series of covalent intermediates that permit the escape of products at appropriate times in the catalytic cycle.

Such systems would seem, in some cases fortuitously, to be attractively simple models for the tightly fixed nature of the two aspartate side chains in the active site of

the enzymes shown in the X-ray crystal structures (see above). All the above studies have found that an anhydride is formed, in some cases transiently, between the two carbonyls as a final step in the intramolecular catalysis. Analogously, it was proposed that an active-site anhydride is formed between the two aspartate side-chains in the last step in the catalysis by the aspartic proteases. This has been shown not to be so by the studies of Antonov[56] on ^{18}O-label incorporation into the active site aspartates. Thus, these types of model studies have contributed little to our understanding of the mechanism of the aspartic proteases.

More recently, model studies have been undertaken on a group of strained bicyclic amides by the group of Brown. These have been proposed as models for peptide hydrolysis by the serine,[137] cysteine[138] and aspartic proteases.[139]



n = 1 or 2: m = 1 or 2

Figure 1.5.3: The strained amides of Brown

In these compounds the nitrogen lone pair is twisted out of the plane of the carbonyl group by varying amounts, increasing with decreasing n and m. The degree of twisting (or the pyramidalisation of the nitrogen) correlates with the rate of hydrolysis by both acid and base, but only to a certain point.[4b] This is intuitively satisfying as it would be expected that twisting of the lone pair out of conjugation with the adjacent carbonyl would decrease amide resonance dramatically (see Section 1.4.1.3). After a certain angle of twist is reached there is no more stabilisation to be lost and so no more rate increase would be observed. Similar results have been obtained by Antonov[140] using quantum mechanical calculations, which showed a linear relationship between the pK_a of the amide nitrogen and its degree of pyramidalisation. The acid-catalysed hydrolysis of amides is controlled by both the rate constant for H_2O attack on the protonated amide ($A\text{-}H^+$) and by the extent of protonation.[141] The greater the distortion the greater the value of the kinetic pK_a for $A\text{-}H^+$. This has been interpreted[142] as providing evidence for water attack on an N-

protonated form. This is in accord with the mechanism proposed by Hemmings, in which protonation occurs on the scissile bond nitrogen prior to nucleophilic attack by a water molecule on the amide carbonyl. The thermodynamic site of protonation is undoubtedly the nitrogen atom, but there could be a "greater degree of involvement of H_2O attack on O rather than N-protonated amides."[141]

Thus these studies also provide support for Pearl's proposal of substrate distortion being a key part of the catalytic armoury of the aspartic proteases. However, once again hydrolysis of these amides is, unsurprisingly, most efficiently catalysed by dicarboxylic acids that can form anhydrides readily.[139b] They would therefore appear to have have little relevance to current thinking on aspartic protease catalysis.

1.6 Introduction: The Human Immunodeficiency Virus-1

1.6.1 HIV-1 and its target cells

The human immunodeficiency virus (HIV) is a pathogenic human retrovirus, carrying its genetic material as a single positive strand of RNA. Two related strains of the virus have been identified, HIV-1 and HIV-2, of which HIV-1 is both more widespread and more pathogenic in man. They are members of the lentivirus subclass of retroviruses (reviewed in[143]). As such they share some similarities with the archetypal lentivirus, the visna virus of sheep *e.g.* it is cytolytic and immune disruptive. They are, however, much more similar to the primate lentiviruses, displaying the same life cycle and mechanisms of toxicity (single cell lysis and syncytium formation).[144] HIV has been identified as the aetiological agent in acquired immune deficiency syndrome, AIDS.[145]

Those cells that express the CD4 molecule ($CD4^+$ cells) on their surface, which binds to the viral coat protein, gp 120,[146] are vulnerable to HIV infection. The function of CD4 will be discussed below. The specificity (tropism) for CD4 is bestowed by various parts of the viral envelope protein.[147] However, small changes in and around the V3 loop of gp120 (the principal neutralising determinant of HIV) can greatly alter the cell tropism.[148] The identification of the receptor for HIV makes it

unique, so far, amongst all human retroviruses. The tropism for $CD4^+$ cells is probably relative rather than absolute as cells that do not express cell surface CD4 nor have CD4 mRNA can be infected with HIV.[149] Also, some T-cells that express CD4 at a high level have been shown to be resistant to HIV infection.[150] CD4 is also expressed by monocytes, colonic cells and glial cells of the brain, making all these cells targets for HIV.

1.6.1.1 Overview of HIV infection

HIV binds to a $CD4^+$ cell, and the viral core enters the cell after fusion of the viral and cellular membranes (mediated by the viral transmembrane protein gp41). The viral genomic single strand RNA is reverse transcribed using viral reverse transcriptase/ribonuclease H to viral double strand DNA, which is then integrated at specific sites in the host cell chromosome by viral integrase. This process only occurs efficiently in activated T-cells.[151] After a period of very low levels of replication (clinical latency) the virus then replicates rapidly, progeny virus bud from the cell and the host cell dies. At this stage the presence of gp120 on the cell surface also causes CD4 bearing cells to aggregate, forming giant, multinucleated syncytia, which results in cell death. The period of clinical latency is usually long (8-10 years) and in this time stable infection can be established in certain cell lines. The trigger to leave the so-called "slow/low" mode and enter the cytopathic "fast/high" mode is not known.

1.6.1.2 Role of CD4 and T-helper cells in immune reponse

T-lymphocytes recognise peptide antigens (an antigen is any molecule that will provoke an immune response) presented on the surface of an antigen presenting cell (APC). This recognition triggers T-cells to carry out their function, either cytotoxic action (T-8 cells) or helper action (T-4 cells). Of these cell types T-4 cells, or helper T-cells, express CD4 on their surface and are therefore targets for HIV infection. Both types of T-cell recognise peptide antigens bound to a polymorphic pocket of one of the two types of major histocompatibility complex proteins, MHC-I and MHC-II.

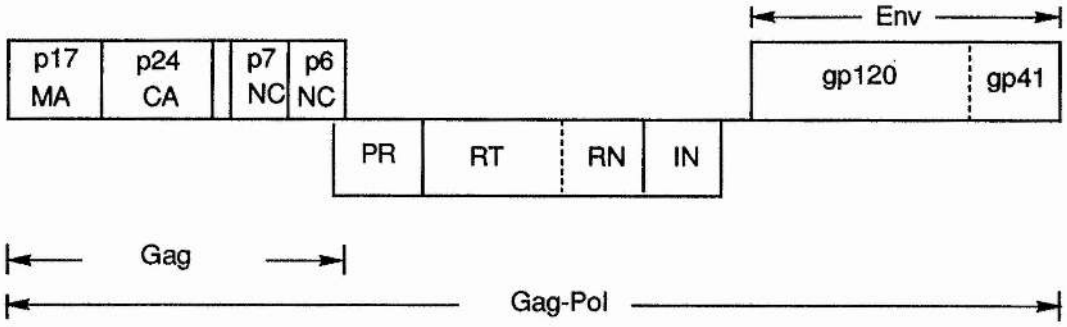
MHC's are a group of proteins involved in self/non-self recognition.

Antigen associated with class II MHC arises from exogenous antigen taken up by macrophages and B-cells (which are APC's) and presented on the cell surface. Class II MHC associated antigen is recognised by helper T-cells. This recognition occurs through interaction of the αβ heterodimeric T-cell receptor (TCR) on the T-cell with MHC II, around the polymorphic antigen binding pocket. Contact between the antigen-MHC complex and the TCR occurs through both the α and β chains of the TCR. Only a small subset of T-4 cells will recognise a given peptide antigen as only a few T-cells will have the correct TCR sequences to recognise the antigen-MHC complex. CD4 and the TCR act as a unit to bind the MHC-antigen complex. The function of CD4 is probably to strengthen the interaction between the TCR on the helper T-cell and the APC by forming subsidiary interactions with class II MHC. The likely site of interaction between CD4 and MHC II is the non-polymorphic region of MHC II. In humans the MHC is known as human leukocyte antigen (HLA).

Cellular immunity (that employing macrophages) and most humoral reponses (those dependent on antibodies and acting against extracellular phases of infection) are T-helper cell dependent. T-cells act by causing B-cell proliferation and differentiation. They also release lymphokines when they are activated, stimulating B-cells and increasing cellular resistance to viral infection. Thus, HIV infection of T-helper cells eliminates the host's ability to mount any kind of defence against a pathogen and so the host succumbs to opportunistic infections. The mechanism by which it does this is unknown, and some possibilities are discussed in Section 1.6.1.5 below.

1.6.1.3 HIV and its life-cycle

Shown below (Figure 1.6.1) is the genomic organisation of the proteins that make up the virus.



where MA = matrix protein
CA = capsid protein
p6/p7 NC = nucleocapsid protein
RN = ribonuclease
PR = protease
RT = reverse transcriptase
IN = integrase

Figure 1.6.1: Genomic organisation of HIV-1

The *gag* polyprotein contains the structural proteins of the virus; the matrix protein, the capsid and the nucleocapsid proteins. The *env* polyprotein contains gp120, the outer coat protein and gp41, a transmembrane protein associated with gp120. The *pol* region contains the enzymes necessary for viral replication, the protease, reverse transcriptase, ribonuclease and integrase.

Given below in Figure 1.6.2 is a representation of the overall structure of HIV-1. The virus possess a bullet-shape core (characteristic of lentiviruses) made up of the capsid (CA, p24) protein. The core contains the 2 single strands of genomic RNA, bound to NC and a few copies of the viral enzymes. The outer protein shell is an icosadeltahedron, made of the matrix protein (MA, p17). The lipid membrane is derived from the host cell that the virus budded from.

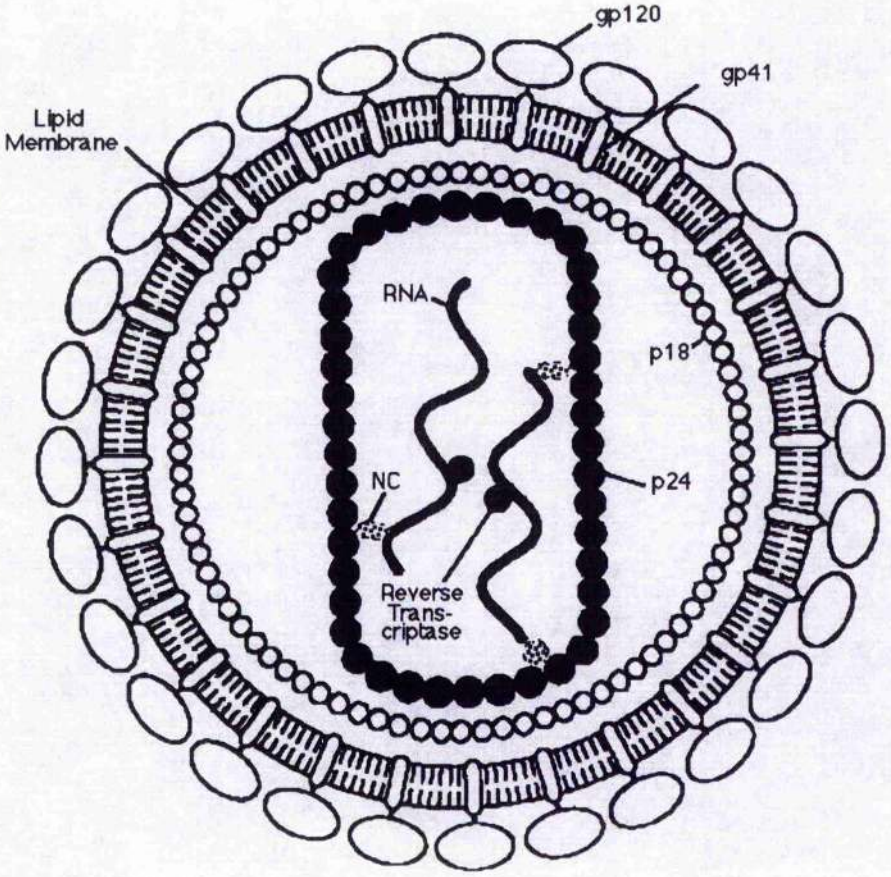


Figure 1.6.2: Overall structure of HIV-1[3]

The genomic organisation of HIV-1 is given below (Figure 1.6.3), showing the different reading frames and splicing patterns. The viral genome is extremely small (just under 10 kb) so production of the nine different viral proteins requires that sections of the genome be utilised more than once. Two reading frames and multiple splicing patterns (4 donor and 6 acceptor sites[152]) of the primary RNA transcripts allows the virus to "re-use" large sections of its genome to code for more than one

protein. The non-spliced transcript provides both the genomic RNA and mRNA for translation into the *gag-pol-env* polyprotein. The transition from the *gag* to the *pol* reading frame is brought about by a -1 ribosomal frameshift that occurs just after the end of the p6 coding region. The shift occurs in around 11% of cases,[153] allowing the production of the 160 kDa *gag-pol* polyprotein. The other 89% of cases produce *gag* alone *i.e.* translation is terminated soon after the frameshift site if the frameshift does not occur. As already described *pol* contains the viral enzymes and *env* the outer coat proteins. The other proteins are regulatory proteins and, along with a number of cellular proteins, govern the progress of the virus through the various stages of infection (see Section 1.6.1.4).

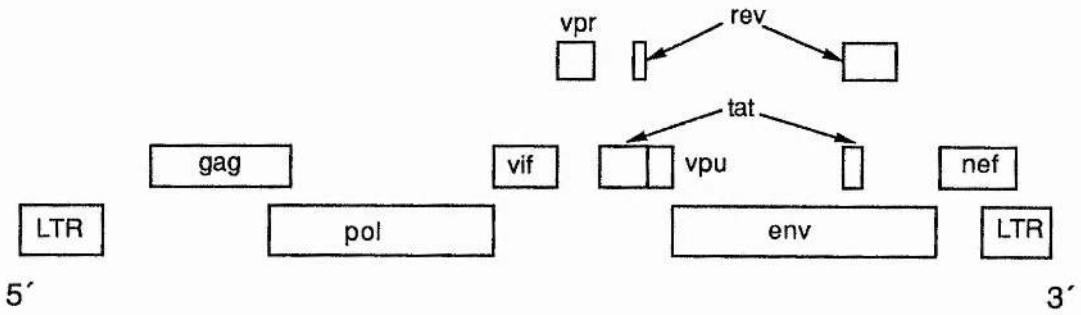


Figure 1.6.3: The HIV-1 genome showing splicing patterns
where LTR = long terminal repeat

1.6.1.4 Viral replication

Given that the structure and the protein constituents of the virus are known it became of interest to study how the various properties of the proteins are coordinated to replicate the virus. A model for the replication cycle has been developed and can be broadly divided into the following steps:

i) Attachment - Viral coat protein gp 120 recognises and binds to cell surface CD4.
ii) Penetration - Viral and cellular membranes fuse, mediated by gp41 and the viral core enters the cell.

iii) Uncoating - Viral RNA and replication enzymes are released into the cell by proteolysis of the viral capsid, probably by the viral protease.

iv) Reverse transcription - Proviral DNA is produced from RNA by viral reverse transcriptase.

v) Integration - Proviral DNA is irreversibly integrated into chromosomal DNA in a site-specific manner by viral integrase. This is only efficient in activated T-cells.

vi) Expression - Viral DNA is transcribed and translated by cellular apparatus to give viral proteins and RNA. The amount and type of proteins produced is tightly regulated.

vii) Post-translational modification - *env* proteins are glycosylated and the N-terminal Gly of p24 is myristoylated by cellular enzymes.

viii) Assembly - Viral proteins migrate to the inside of the plasma membrane, proteolytic processing begins and the virion core is assembled.

ix) Budding - Virion core buds out from the plasma membrane, removing some of it as it does so and proteolytic processing is completed.

In the early stages of infection short (≈2 kb), multiply spliced mRNA's are produced, coding for the regulatory proteins. Later in infection full-length mRNA's are produced coding for the *gag* and *pol* polyproteins. The trigger for this change is unknown (see below). The crucial stage in replicating the virus occurs in step (v) to (vi) above *i.e.* the transition from stably integrated viral DNA producing a small subset of viral proteins to actively transcribed DNA producing new virus. This transition is controlled by a complex web of interactions between a variety of cellular and viral regulatory proteins. The 5′ and 3′ long terminal repeats (LTR's) contain many *cis*-acting regulatory elements that bind a wide range of viral and host proteins. The regulatory proteins tat, rev and nef control the level of synthesis of viral proteins, including each other. The trans-activating protein, tat, binds to the so-called TAR sequence in the R region of the 5′ LTR and enhances transcription therefrom, causing a generalised increase in the level of viral proteins.[154] It is also implicated in increasing the efficiency of mRNA translation, possibly at the level of mRNA transport or stabilisation.

The regulator of virion proteins, the rev protein, acts to differentially affect the synthesis of viral proteins, no structural proteins being synthesised in its absence.[155] It has two *cis*-acting target sequences in the *gag-pol* and *env* genes, the repressive CRS sequence and the activating CAR sequence. It is thought that the CRS

sequence confines the mRNA to the nucleus, but that the CAR-rev interaction can override this.[156] The negative regulator, nef, lowers expression of all viral genes by acting on a sequence in the LTR.[157] The function of the vpr and vpu proteins is not well understood and vif is thought to be involved in viral infectivity.[158] Deletion of any of these three proteins had no observed effect on viral infectivity or pathogenicity.[159]

These proteins act, together with several cellular proteins (such as NF-κB[160]), to regulate entry of the virus into the late stage of the replication cycle (stage (vi) above) and become active. It was previously thought that the virus was in a so-called latent stage for the early part of infection and that there was a cellular trigger for entry to the late stage. However, it has recently been shown[161] that HIV is never latent, merely replicating at a previously undetectable level. Thus, the cycle given above proceeds continously after integration of viral DNA into the host chromosome. However, infection only results in a clinically observable change after some considerable period of time as very low levels of virus are produced in the early stages of the disease. It has been observed that the nature of the virus changes in the course of infection. At the early (latent) stages of the disease the virus is slow replicating, low yielding and does not cause syncytia. In the late stages the virus replicates rapidly and with high yield, readily establishes productive infection and causes syncytia formation.

It has been postulated that the long duration of the early stage is due to immune surveillance suppressing viral replication, during which time the virus mutates rapidly, eventually exceeding the host's ability to suppress replication. These viral escape mutants (mostly mutants in the V3 loop of gp120) are then extremely diverse[162] constituting several quasi-species. It may be that this continous antigenic drift, arising from the high mutation rate, and antigen specific triggering of different T-cell clones leads to exhaustion of containment and immune collapse.

The extremely high mutation rate in HIV and other RNA viruses is partly due to the lack of exonucleolytic (proofreading) activity in the viral reverse transcriptase.[163] This results in extremely error-prone replication (around 1 mutation per 10,000 bases or 1 mutation per virus per generation).[164] However, the possession of a high fidelity reverse transcriptase is of little use to the virus as transcription of the integrated viral DNA is carried out by the rather error-prone cellular RNA polymerase II.

1.6.1.5 Mechanisms of pathogenicity

The results of HIV infection of the immune system are:

i) Severe depletion of $CD4^+$ cells.

ii) Loss of T-helper cell function (independent of the decline in total $CD4^+$ cell count).

Any of the stages in the viral life-cycle could be cytotoxic, but the most likely are stages (i), (vi) and (ix). Stage (i), initial binding, may disrupt the proper functioning of the CD4-MHC II-TCR complex (see below). This complex is most important in the correct presentation of antigens to B cells and thus the mounting of an antibody-mediated immune response. Stage (vi), expression of viral regulatory genes may disrupt cellular transcription or translation. The load on limiting cellular machinery may be too high to allow the cell to survive *i.e.* viral mRNA competes for limited protein synthesising resources to the detriment of the cell. Stage (ix), budding, may destroy membrane cohesion and cause the membrane to become punctured. However, due to the close parallels between graft versus host disease and the early stages of HIV infection it has been proposed that the pathogenicity of the virus is due to its ability to trigger an autoimmune response against helper T-cells.

A type of autoimmune disease may arise through an immune response to gp 120, particularly the CD4 binding site. Anti-idiotypic (antibodies against antibodies) gp120 antibodies might bind to CD4 much as gp 120 is seen to do *in vitro* and thereby block the function of $CD4^+$ T-helper cells.[165,166] This binding of antibodies to $CD4^+$ cells will also lead to their complement mediated destruction and programmed cell death (apoptosis).[167] It will also make them a target for the normal antibody targeted cell killing.

It has recently been proposed that gp 120 is a so-called superantigen and it is this property that gives the virus its powerful immune-disruptive effect.[168] Superantigens bind to class II MHC and the β chains of the TCR in a non-β chain specific manner and stimulate a powerful proliferative response in the T-cells bearing the β chains recognised by the superantigen. This activation of T-lymphocytes bearing certain Vβ sequences as part of the β chain of the TCR may result in clonal deletion of these T-lymphocytes. It is by such a mechanism that the immune system deletes T-cells specific for self antigens.[169] Thus, it is proposed that gp120 causes constant,

inappropriate activation of T-cells bearing the correct receptor (CD4) by binding to both TCR and CD4 at the same time. Experimental support for this theory has come from recent work[170] demonstrating for the first time the existence of a human superantigen. It was found that the nucleocapsid protein from rabies virus is a superantigen specific for Vβ8 bearing T-lymphocytes *i.e.* T-cells having the so-called V8 sequence in the β chain of the TCR.

Another proposal is based on the close structural resemblance between a model for gp120 and the C-terminus of HLA-A2 α_2 chain.[171] Thus, it is hypothesised that gp120 on the surface of infected cells, in association with CD4, will induce an immune response. The antibodies so induced will target HLA-A2 α_2 chain bearing cells *i.e.* T-helper cells, as well as gp120 bearing cells, resulting in their elimination. This is consistent with the increase in activity of $CD8^+$ killer T-cells and the decline in the $CD4^+$ cell count in the early stages of HIV infection as the $CD8^+$ cells are engaged in killing the $CD4^+$ cells. In either case the decline of the T-helper cell population means that most B and T-8 cell responses cannot be mounted against a pathogen.[172]

1.6.2 The HIV-1 protease

After it was recognised that HIV contained a protease that was absolutely required for viral maturation a huge multi-disciplinary effort was launched to study the enzyme, its properties and its place in the viral life-cycle. The state of knowledge on the HIV protease is summarised in the following sections.

1.6.2.1 Role of the protease

The HIV-1 protease is, like the other viral enzymes, a minimal enzyme, just large enough to carry out its allotted task. Its homodimeric nature (see below) allows the use of a very small coding region whilst retaining activity. Protease activity is mostly expressed at a late stage of the life-cycle (see Sections 1.6.1.3 and 4), when it releases and activates the viral structural proteins and enzymes from the viral polyprotein to form mature infectious virus. The cleavages carried out are involved

and facilitate the assembly of complex arrays of proteins for which self-assembly may not be adequate. It has recently been shown that the protease also has some role in the early stages of the life-cycle; at the integration stage[173] and the activation of proviral transcription.[174] Protease activity also appears to be involved in the uncoating stage as protease defective virus particles are stable under conditions that disrupt mature virions.[175] The protease of the related retrovirus, equine infectious anaemia virus (EIAV) has been shown to be involved in hydrolysis of the nucleocapsid (NC) proteins to release viral RNA into the cell.[176] Very recently it has been shown that the HIV-1 protease has a similar role. A second protein processing pathway, involving regulated *in situ* cleavage of the NC protein within intact capsids, has been found.[177] Thus, the protease is required for early events in replication and protease inhibitors may therefore be useful for preventative therapy, as well as for treatment to prevent viral spread in the late stages of infection.

During steps (vii) and (viii) in the replication cycle above, the *gag* and *gag-pol* polyproteins are myristoylated, by a cellular myristoyl transferase,[178] on glycine-1 of the matrix protein MA.[179] This encourages migration of the polyproteins to the inside of the cell membrane,[180] where the myristate hydrocarbon chain embeds itself in the membrane, anchoring the polyproteins. This decrease in available dimensions for movement increases the effective concentration of the polyproteins and thus promotes dimerisation to form active protease. Once dimerisation occurs release of the protease from the polyprotein is autocatalytic.

The kinetics of proteolytic processing show that the processing occurs in *trans*[181] *i.e.* intermolecularly, which is the least sterically demanding route. It is possible that the so-called p6* region (which lies between the coding regions for p6 NC and the protease) of the *gag* polyprotein sterically blocks the active site in the same way as the active site of the aspartic protease zymogens is blocked by the propeptide.[94] However, there is a poor sequence alignment between p6* and the propart of the aspartic proteases, so they may not function in the same way. Instead, the protease may be activated by displacement of p6* on conformational change induced by association of the *gag-pol* polyprotein with *gag* or viral RNA or by pH change as the immature virion buds from the cell.[182] On deletion of the p6* region from the viral genome the efficiency of *gag-pol* processing by the protease is increased[183] so it is

possible that p6* is involved in some way in reducing or regulating protease activity. The importance of the timing of processing is shown by the loss of infectivity and viral assembly in cells infected with HIV-1 that bears the protease as a single chain monomer *i.e.* pre-assembled and permanently active.[184]

The cleavages carried out by the protease in the viral polyprotein are shown in the diagram below, Figure 1.6.4.

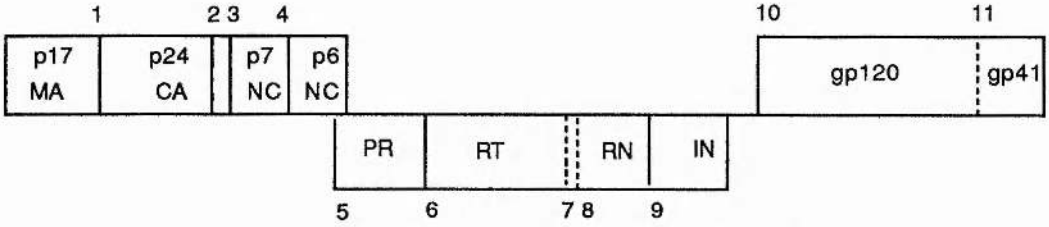


Figure 1.6.4: The in vivo cleavage sites of the HIV-1 protease

The various cleavage sites in the polyprotein are not acted upon by the protease at the same rate, resulting in an ordered release of the constituent proteins.[185] This ordered release was shown to correlate with the rate of cleavage of synthetic peptides based on these *gag* cleavage sequences.[186] Ordered cleavage may be required for regulation of proper interaction and assembly of the viral core and coat proteins.[187]

1.	-Ser-Gln-Asn-Tyr	*	Pro-Ile-Val-Gln-
2.	-Ala-Arg-Val-Leu	*	Ala-Glu-Ala-Met-
3.	-Ala-Thr-Ile-Met	*	Met-Gln-Arg-Gly-
4.	-Pro-Gly-Asn-Phe	*	Leu-Gln-Ser-Arg-
5.	-Ser-Phe-Asn-Phe	*	Pro-Gln-Ile-Thr-
6.	-Thr-Leu-Asn-Phe	*	Pro-Ile-Ser-Pro-
7.	Leu-Glu-Lys-Glu	*	Pro-Ile-Val-Gly-
8.	-Ala-Glu-Thr-Phe	*	Tyr-Val-Asp-Gly-
9.	-Arg-Lys-Ile-Leu	*	Phe-Leu-Asp-Gly-
10.§	-Arg-Glu-Lys-Arg	*	Ala-Val-Gly-Ile-

§ Note: This cleavage is probably carried out by a cellular tryptase.[188]

Figure 1.6.5: Sequences of cleavage sites in the gag-pol polyprotein

1.6.2.2 Classification of the protease

Having identified several areas where protease activity is crucial for viral replication it then became of interest to identify the class of protease to which the HIV-1 protease belongs and thereby rationally develop inhibitors for it. The protease was classified as an aspartic protease initially on the basis of its sequence,[189] as it exhibits the absolutely conserved active site triplet of aspartic proteases (Asp-Thr-Gly) (see Section 1.2.2). The sequence shows that there is only one aspartate residue per chain, so to provide the pair needed for catalysis the enzyme must be a homodimer, with one aspartate being provided by each 99 amino acid monomer (Asp-25). This makes the enzyme exactly symmetrical, except for the differing protonation states of the aspartates *i.e.* it possess almost perfect C_2 symmetry. This symmetry is broken on substrate or inhibitor binding,[190] see below. The classification as an aspartic protease was supported by inhibition by pepstatin A,[191] the archetypal aspartic protease inhibitor (see Section 1.1) and by abolition of activity on mutation of the active site aspartate to alanine.[192] Definitive proof of the classification was provided by the crystal structure[193] which clearly showed the expected dimer structure and the active site containing the D-T-G triads.

1.6.2.3 Properties of the protease

The protease is a somewhat atypical aspartic protease in many respects. Its pH optimum is unusually high, around 5.5,[194] making it more like renin than the rest of the aspartic protease family. The reason for this rather high optimum compared to most of the monomeric aspartic proteases is not clear (see Section 1.2.3). The protease is homodimeric, with a monomer molecular weight of around 10,750 Da[194] and the dissociation constant for dimer formation is 10 nM.[195] The dimerisation allows the functional structure of the larger, monomeric aspartic proteases to be conserved whilst using considerably smaller polypeptides. It is thus a minimalist aspartic protease. The necessity of dimerisation for activity may provide a control point in the maturation of the virus (see Section 1.6.2.1)[196] preventing premature processing.

It also seems that the retroviral proteases are somewhat less catalytically active

than the mammalian aspartic proteases, with the highest turnover number for HIV-1 protease[194] around 70 s^{-1}, whereas for pepsin the turnover number can be as high as around 400 s^{-1}.[197] This low efficiency may not be a limitation on viral maturation as the ratio of protease molecules to cleavage sites is around 1:80.[198]

1.6.2.4 Substrate specificity

The *in vivo* cleavage sites of the protease show that it is required to perform a small number of highly specific cleavages at sites containing a wide range of amino acid sequences. This small number of cleavages in a very large protein provides a false impression of the specificity of the enzyme as it shows good activity against a very diverse range of peptides. There are three factors that are involved in determining the rate at which a given site in a peptide or protein substrate will be cleaved; accessibility, conformation and sequence. These three factors must combine in the *in vivo* situation to give the low number of cleavages observed in the *gag-pol* polyprotein. Unfortunately, the predictive value of these criteria is sometimes poor.[199]

In protein substrates of the HIV-1 PR there is an absolute requirement for structure, as denatured *gag* polyprotein and other denatured proteins are not substrates.[200] Not only is some kind of structure required it appears that the conformation of the peptide or protein has a profound effect on cleavage efficiency[201] *i.e.* a particular kind of structure is required. This is supported by the observation that heat-denatured rat lactate dehydrogenase is not a substrate for the protease, but partially pH denatured rat LDH is.[199] However, 2D-NMR investigations of peptide substrates for HIV-1 PR and AMV PR have shown that they have only random conformation in solution.[202] Conformationally restrained circularised peptides have been shown not to be substrates.[201] It has also been found that a decameric peptide from the active site of LDH is cleaved as quickly as the same sequence in the protein. However, the conformations of the peptide free in solution and in the enzyme tertiary structure are very different.[199]

The third determinant for cleavage, sequence, has been extensively investigated using small synthetic peptides, usually based on the *gag-pol* cleavage sites shown

above. This work is based on the assumption that cleavage sites in proteins and peptides are often cleaved at similar rates.[203] These peptides usually occupied S_4-S_4'. In some sequences it has been found that seven amino acids, spanning P_4-P_3', are required for cleavage,[204] while some sequences do not require a P_4 residue to be cleaved.[205] Therefore, some sequences require only 6 residues to be cleaved, other sequences require 7 residues.

Three of the *gag-pol* sites contain the consensus sequence Ser(Thr)-Xaa-Yaa-Aromatic*Pro, which is a common cleavage sequence amongst the retroviral proteases.[206] The remainder show hydrophobic residues at P_1 and P_1' which also occur frequently in retroviral protease cleavage sites,[207] with no ß-branched residues allowed at P_1 in either protein or short peptide substrates.[208] This could be related to the observation that *cis* proline residues (as are found in several HIV-1 PR cleavage sites) are almost never preceded by ß-branched amino acids.[209] Studies aimed at determining which of sequence and conformation is the dominant influence on the rate of cleavage have provided ambiguous results.[203]

An extensive analysis of retroviral protease processing sites has allowed the sites to be classified into classes;[210] type 1 sites have Pro at P_1' and type 2 have Ala, Leu or Val at P_1'. A similar classification[211] postulates a third class having Glu/Gln at the P_2' position and hydrophobic residues at P_1 and P_1'. It is possible that these sites are functionally different and contribute toward the ordered processing discussed above. In this context it is interesting to note that the amino terminus of the capsid protein CA is always a type 1 site and the carboxyl terminus is always a type 2 site, but this is the only example of such conservation in cleavage sites in 10 different viruses.[202] Also, sites cleaved in non-viral proteins are nearly always of Class III.[210]

In the course of these investigations it has been observed that k_{cat} for cleavage increases with length of the peptide, while K_M remains approximately constant[211] as was observed in the cleavage of peptides by the monomeric AP's (see Section 1.2.3). However, this was not observed in a similar study by Billich *et al.*[212] The reason for this discrepancy is as yet unclear. If it is shown that there is no change in k_{cat} with peptide length then the mechanism of HIV-PR activation of peptide substrates will need to be reexamined.

Extensive experimentation and analysis has allowed the determination of the

favoured residues to occupy each of the subsites P_4-P_4' in the enzyme. Analysis of 40 cleavage sites in various (viral and non-viral) proteins allowed Poorman *et al.*[213] to assign specificity indices for each of the subsites in the enzyme binding site. A similar analysis of cleavage of synthetic peptides[205] found that the S_1 site was much more stringent in its requirements that the S_1' site. Both studies also found that the S_2' site was by far the most stringent, followed by S_1 and S_2. These observations are difficult to rationalise given the symmetry of the protease, which means that the S_n and S_n' subsites are made up of the same residues. The explanation probably lies in cooperativity in substrate side-chain binding between the various subsites.

Context dependency has also been shown to be an important influence on the efficiency of cleavage of a given peptide bond.[214] This study compared the efficiency of cleavage of p1-p7 (Met-Met) and p17-p24 (Tyr-Pro) junctions when the sequences flanking the scissile bond are interchanged *i.e.* increasing amounts of one site replaced by the other. The effect of a given residue on the kinetic parameters for cleavage varies greatly with the residues that are around it in the substrate. The influence of residues distant from the scissile bond is further demonstrated by the large difference in the rate of cleavage of two peptides, of overall different sequence, at the sequence Asn-Phe*Pro. The peptides were found to be cleaved at rates differing by about 30-fold.[182]

Sites very distant from the scissile bond *i.e.* beyond S_4 and S_4' also have a large influence on the rate of cleavage. It has been shown that octapeptides of the same sequence as the cleavage sites in the polyprotein do not show the large differences in cleavage rates that are shown by those sites in the *gag-pol* polyprotein.[214] Therefore, additional conformational or steric features or binding determinants beyond P_4 and P_4' are required by the protease to achieve maximal rates of cleavage. This is curious as the crystal structures of the protease complexed with various inhibitors clearly show that residues beyond P_3 and P_3' have very little, if any, specific interaction with the enzyme.

1.6.2.5 Mechanism of the protease

The mechanism of the HIV-1 protease has not been explored with the same

thoroughness as that of the monomeric proteases as it is assumed that they will be the same. The studies on the kinetics of product release have shown that with some substrates the N-terminal fragment is released first, as was seen for several pepsin substrates (see Section 1.3.1.1). However, some substrates show random release of the two fragments from a central complex.[215] It is not known if the ordered release is a pH dependent phenomenon, as it is for some pepsin substrates (see Section 1.3.1.1). The studies of Northrop *et al.*[110] on solvent isotope effects on pepsin catalysis also support an ordered release mechanism.

Studies[215] on ^{18}O-label incorporation from ^{18}O-labelled water into peptide substrates have indicated that the intermediate in peptide bond hydrolysis is a neutral amide hydrate,as suggested by Blundell for the monomeric proteases (see Section 1.4). Studies on the pH-rate profile[216] have shown that at high pH the rate determining step is chemical, whereas at low pH it is product release (or conformational change). Solvent isotope effect studies have shown that there is no effect on $^{D}V/K_m$ and a variable, but greater than 1, effect on V_{max}. The lack of an effect on $^{D}V/K_m$ is ascribable to a high forward commitment and the effect on ^{D}V arises from the simultaneous transfer of two protons in the breakdown of the tetrahedral intermediate.[216] These effects are consistent with those observed with the monomeric aspartic proteases (see Section 1.3.5).

Chemically, therefore, the mechanism of the HIV-1 protease is essentially identical with that of the monomeric aspartic proteases. Kinetically, the mechanism is less well understood.

1.6.3 Inhibition of the HIV-1 protease

The HIV-1 PR is a promising therapeutic target in AIDS and HIV infection for several reasons:

i) Its central role in the life-cycle. It has been shown that mutagenic inactivation (see above) or deletion[217] of the protease results in the production of non-infectious, immature virions. Unlike reverse transcriptase inhibitors such as AZT[218] and ddI, protease inhibitors are active against virus in chronically infected cells as well as attenuating acute infection of uninfected cells.[219] This is not unexpected given that

the protease acts mostly in the latest stages of the viral life-cycle. Protease inhibitors have been shown to prevent viral replication in chronically infected cells and to prevent viral spread to uninfected cells.[220] Other inhibitors have been shown to be able to clear viral signs from acutely infected cells in a few days.[221] This action may derive from the inhibition of the action of the protease on the viral nucleocapsid protein early in the replication cycle (see Section 1.6.2.1).

ii) High selectivity over host enzymes can easily be achieved. The unusual cleavage sites for the protease (aromatic-proline), which are not cleaved by mammalian proteases, allow highly selective inhibitors to be designed based on an aromatic-proline scissile bond analogue. This will make them much less cytotoxic than reverse transcriptase inhibitors, a significant advance as AZT is very toxic in humans.[222]

iii) The low rate of mutation in the protease. A further problem with AZT and related therapies is the rapid emergence of resistant mutants.[223] These result from single point mutations in the reverse transcriptase sequence. The likelihood of protease inhibitor resistant mutants emerging is much smaller as the protease is a very small and highly conserved enzyme, leaving little opportunity for active mutants to occur. This is supported by recent observations[224] that the protease gene from a number of clinical isolates has a very similar sequence to that of the laboratory strain HIV-1 IIIb. The substitutions all occur away from the active site region, suggesting that these enzymes should all be inhibited similarly to, if not identically with, the enzyme from the laboratory strain. There have, however, been reports of the emergence of inhibitor resistant mutants.[225]

1.6.3.1 Development of HIV-1 protease inhibitors

When the HIV-1 protease was identified as a promising site of intervention in the viral life-cycle much of the early work was based on the principles of inhibitor design developed during the search for effective renin inhibitors.[226] Knowledge of the substrate specificity gained from good peptide substrates was used to generate lead peptide sequences. The scissile amide bond in these sequences was replaced by a suitable transition state analogue.[227] In the renin case the sequence was derived from the sequence of angiotensinogen and replacement of the scissile bond in peptides derived from angiotensinogen gave potent renin inhibitors.[228] A similar

approach with the HIV protease using a variety of scissile bond analogues also yielded some very potent inhibitors. In this case the peptides were derived from the sequences of the favoured cleavage sites in the *gag-pol* polyprotein.

Amongst the scissile bond replacements used were phosphinic acid,[229] hydroxyethylamine,[230] α,α difluoroketone,[231] reduced peptide,[232] hydroxyethylene,[233] statine and derivatives,[234] dihydroxyethylene,[235] allophenylnorstatine[236] and phosphonamidate[237] (see Figure 1.6.6). Generally the potency of the analogues increases in the order: reduced peptide, statine, phosphinic acid, α,α-difluoroketone, dihydroxyethylene, hydroxyethylene and hydroxyethylamine.

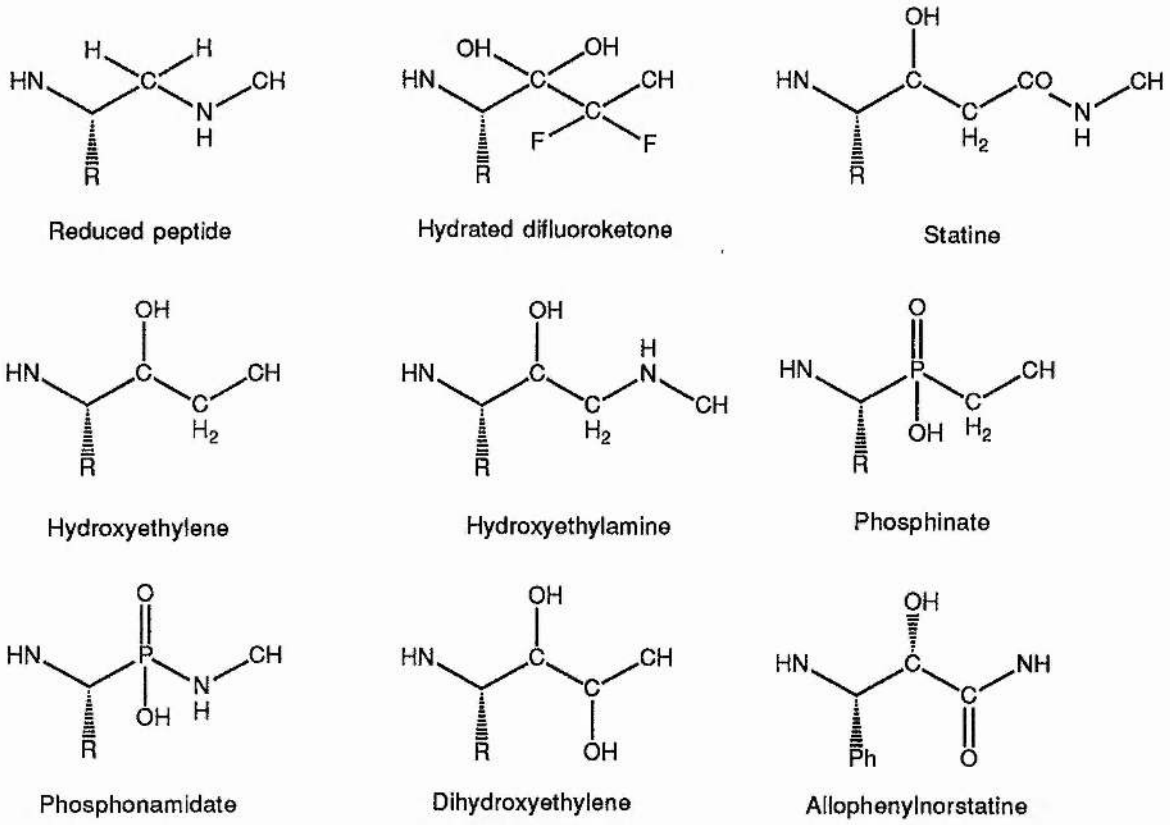


Figure 1.6.6: Scissile bond analogues

All of the inhibitors shown above probably act as so-called collected substrate inhibitors[238] *i.e.* they replace both of the substrates in amide bond hydrolysis, the peptide and the lytic water molecule. Other scissile bond analogues have been developed, but have been less throughly explored *e.g.* amino hemiketals.[239] It has been noted that a substrate of less than 6 amino acids is not cleaved by the

enzyme.[240] However, the same requirement does not seem to apply to inhibitors as good inhibition has been acheived using tetrapeptide mimics (*i.e.* inhibitors that only occupy S_2 to S_2')[233h] and pentapeptide mimics.[233g,i] One of these pentapeptide mimics[233i] is the most potent HIV-1 protease inhibitor so far produced with a K_i of 0.03 nM (Figure 1.6.7).

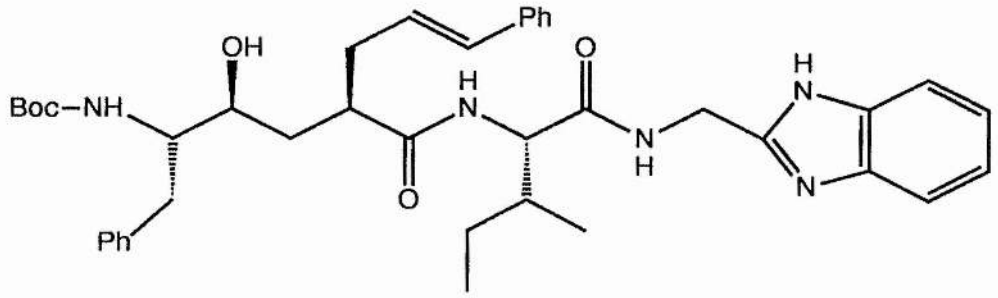


Figure 1.6.7: A pentapeptide mimic inhibitor for the HIV-1 protease

1.6.3.2 Symmetric inhibitors

One of the most prominent differences between the mammalian and the retroviral aspartic proteases is the homodimeric nature of the retroviral enzymes and their resultant symmetry. Recently, inhibitors that possess greater or lesser amounts of symmetry about the scissile bond analogue have been prepared and tested, to exploit this inherent symmetry in the target enzyme. They have been based on hydroxyethylene,[233c,g,h,i] dihydroxyethylene[235c,d], difluoroketone[231a] and phosphinate[241] scissile bond analogues. However, not all the symmetric inhibitors have proved to be improvements over the non-symmetric version.[233d] Symmetric inhibitors offer two pharmacological advantages over the non-symmetric versions. They are somewhat smaller than the unsymmetrical inhibitors, which usually gives higher solubility and better transport across membranes, giving higher *in vivo* efficacy. They also have fewer peptide bonds due to the use of peptide mimicking moieties at their termini, which prolongs their lifetime *in vivo* as they are not such good substrates for host proteases. Many of the peptide-based inhibitors suffer from poor *in vivo* performance compared to their *in vitro* efficacy due to poor absorption into cells and their short lifetime in the circulation. Symmetric inhibitors such as the

one shown in Figure 1.6.8 below overcome these problems to some extent and this inhibitor has been found to be a potent inhibitor of the protease in an *in vitro* and an *in vivo* situation (IC_{50} 0.67 nM *in vitro*).

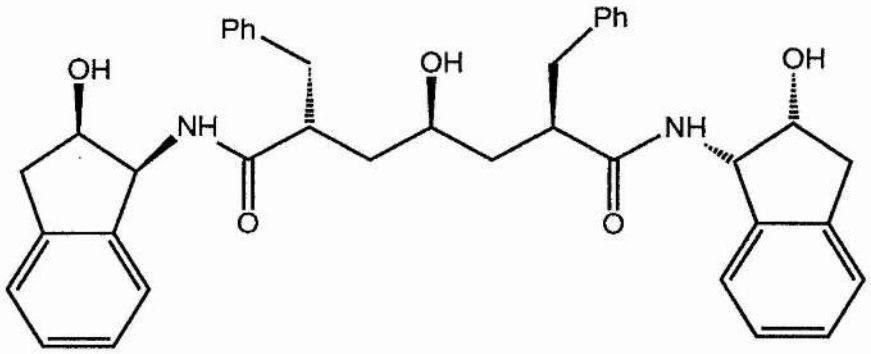


Figure 1.6.8: L-700,417, a potent symmetrical inhibitor[242] *for the HIV-1 protease*

1.6.3.3 Non-substrate based inhibitors

Other routes have been taken to avoid the presence of peptide bonds in HIV PR inhibitors. A screening program at Glaxo U.K. has discovered penicillin-based inhibitors containing almost no peptide like structure.[243] Similar approaches from other groups have unearthed a wide range of compounds that are inhibitors of the protease; disulfonates,[244] dicarboxylic acids,[245] flavones,[246] boronated porphyrins[247] and cerulenin.[248] These compounds are all assumed to bind in the active site, but this has only been demonstrated for the penicillin-based inhibitors. Some metal ions have been found to be non-competitive inhibitors of protease.[249]

A complementary approach to active-site directed inhibition has been prevention of dimerisation by disruption of the dimer interface with peptides derived from the sequence at the dimer interface. This is a promising approach if non-peptide dimerisation inhibitors can be developed as no host protease can possibly be inhibited by this route as they are all catalytically active as monomers. However, to date only moderate (micromolar) inhibition has been achieved by this route.[250]

A search of the Cambridge Small Structures Database for compounds complementary to the protease active site yielded the anaesthetic haloperidol as a lead inhibitor.[251] However, it has since been shown that, although haloperidol is an inhibitor, it is non-competitive and binds additively to the protease with the active site directed inhibitor, pepstatin A (Figure 1.0.1).[252] Thus, in spite of careful design, haloperidol is not an active site directed inhibitor, as expected, but inhibits by an undesigned and unanticipated route, probably by binding to the dimer interface.

1.6.4 Structure of the HIV-1 protease

As mentioned above definitive proof of the classification of the HIV-1 protease as an aspartic protease was provided by an X-ray structure.[193] There have been three other structures of the unliganded protease published, one of low quality using recombinant protease,[253] another at higher resolution also using recombinant protease[254] and one of a chemically synthesised protease.[255] The differences between these last two structures are small.

The structures all show the dimeric structure predicted, with the dimer interface at residues 1-5 and 95-99 of each monomer. The interface consists of two interdigitating β-sheets made up of these residues. In the unliganded state the flaps (residues 33-59) project out into the solvent and are disordered. The flaps serve the same purpose in the HIV-1 protease as they do in the monomeric aspartic proteases. Thus, the flaps exclude solvent from the active site and they form further interactions between the substrate and the enzyme. These conclusions are supported by molecular dynamics simulations, both *in vacuo*[256] and under PBC solvation conditions.[257] The simulations show correlated motions between the flap and another region of the protease, the so-called cantilever, residues 59-75. These areas are correlated both between and within monomers, implying cooperative movements in these parts of the protease as the flaps open and close.

1.6.4.1 Structures of the protease complexed with inhibitors

Of great importance in inhibitor design has been the X-ray crystal structures of the protease complexed with a variety of different compounds. These include inhibitors based on such scissile bond analogues as hydroxyethylene,[258] statine,[259] hydroxyethylamine (JG-365[260] and U-85548e[261]), symmetric hydroxyethylene (A-74704[262] and L-700,411[252]), reduced peptide (MVT-101[263]) and dihydroxyethylene.[264] These structures have provided the drug designer with a wealth of useful information on the nature and size of the enzyme's subsites. The fold of the protease when liganded with an inhibitor (not shown) is illustrated on the ribbon diagram overleaf (Figure 1.6.9).

When the various structures are compared it is interesting to note that parts of the enzyme may shift by up to 1Å r.m.s. displacement on inhibitor binding.[242] This fits well with the information on cooperativity of motion within the enzyme revealed by molecular dynamics studies.[256,257] However, when the various inhibitors are compared they differ very little in their conformation, particularly in the P_1-P_1' region.[265] The constancy of the bound conformation of the published inhibitors has allowed the development of a model for the interaction between our inhibitors and the enzyme (see Section 2.5.1).

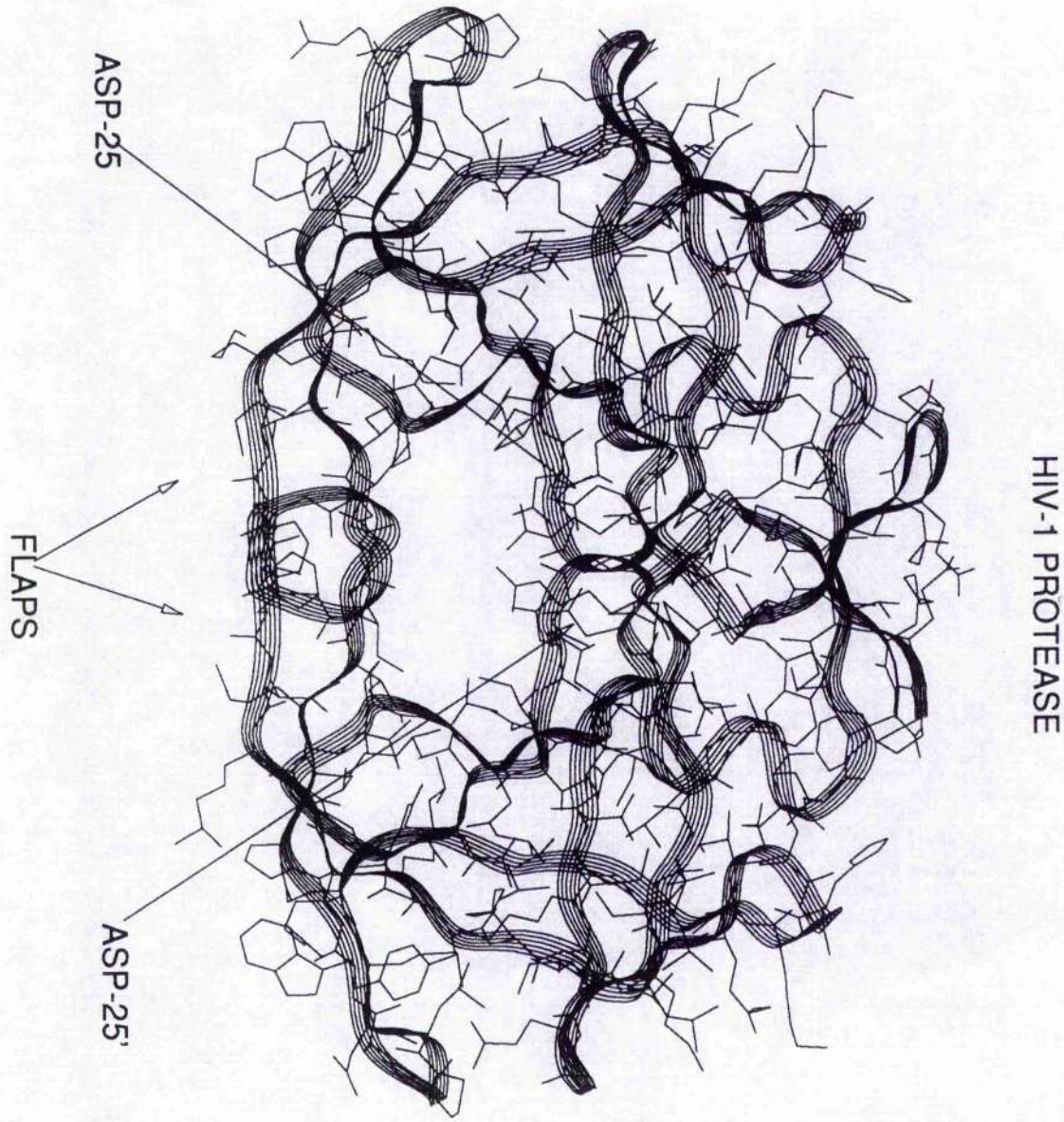


Figure 1.6.9: Overall fold of liganded HIV-1 protease

A most striking feature of these crystal structures is the presence of a water molecule in the active site that acts as a bridge between the flaps and the inhibitor. This interaction is absent in the monomeric aspartic proteases and may present another route to inhibitors of the protease. This interaction is considered to be crucial for some types of inhibitors to achieve optimal binding. A diagrammatic representation of the hydrogen bond pattern around an inhibitor at the active site is shown below (Figure 1.6.10).

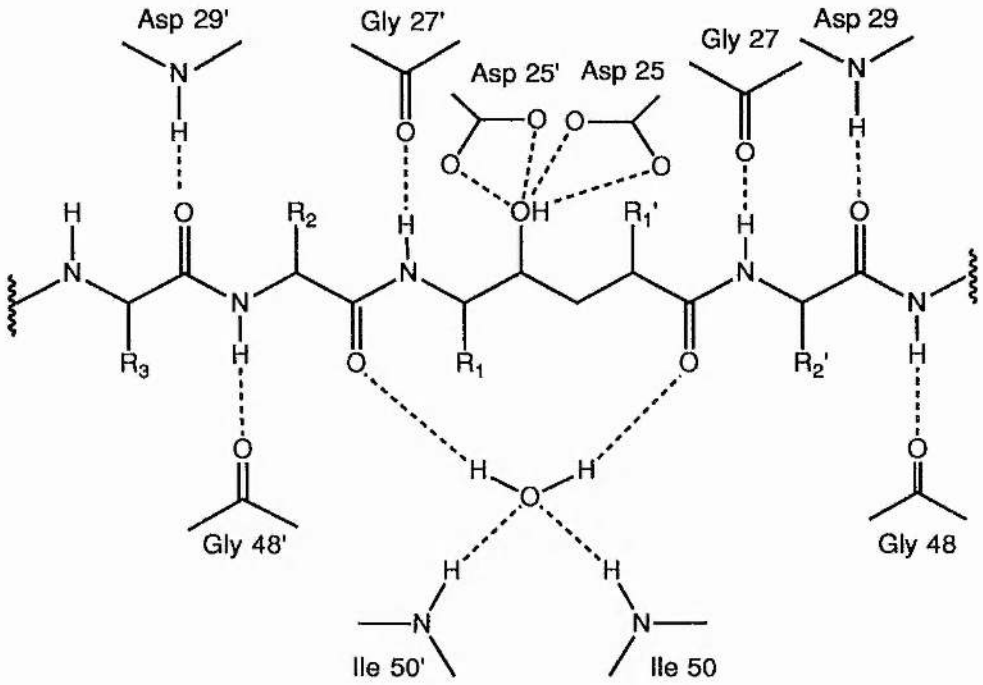


Figure 1.6.10: Hydrogen bonding interactions between the HIV-1 protease and a hydroxyethylene inhibitor[266]

Comparison with the monomeric proteases shows that the P_3 pocket has the Asp 29' residue as its primary binding residue. It has been suggested that this residue does not play the same role in binding substrates as the equivalent residue in the monomeric versions, Thr 219, does.[267] This is because there can only be one hydrogen bond formed between the substrate and Asp-29, compared to two with threonine. Thus, the pivotal role of the P_3 pocket in the monomeric aspartic protease catalysis (see Section 1.4.2) is not mirrored by the HIV-1 protease. For a fuller examination of the importance of the S_3 site of the HIV protease in inhibitor binding see Section 2.5.

There are indications in these structures that asymmetry is induced in the enzyme on inhibitor binding. The amide hydrogen of Gly 51A hydrogen bonds to the carbonyl of Ile 50B, but the amide hydrogen of Gly 51B hydrogen bonds to solvent water.[259,263,82] This is reflected in large differences in the backbone torsion angles for Gly 50A (ψ = -20° to -45°) and Ile 51A (ϕ = -60° to -95°) compared to Gly 50B (ψ = 120° to 160°) and Ile 51B (ϕ = 75° to 120°).[264] Another interesting feature of these structures is that they show some of the inhibitors adopting multiple binding modes.[258d,259,263] This is probably due to the symmetry of the enzyme providing two almost energetically equivalent sets of interactions within the active site, thus allowing two binding modes.

2.0 Results and discussion

2.1 Mechanistic studies on pepsin

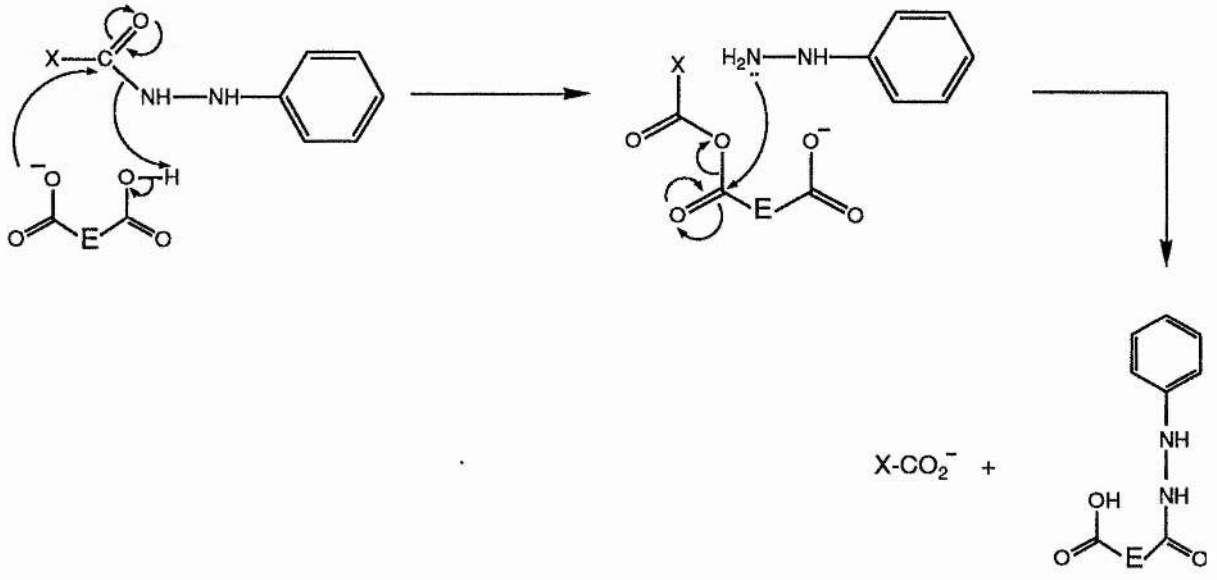
When these studies began the question of the mechanism of the aspartic proteases was still unresolved, with several mutually incompatible theories based on the nucleophilic water concept being propounded. Examination of the early literature showed that there were many questions left unanswered by the nucleophilic water mechanisms and so attention was also turned to the covalent mechanisms (see Section 1.3.1). In an effort to provide more definitive data two parallel sets of experiments were devised, one set based upon the existence of an acyl intermediate and the other based upon the nucleophilic water mechanism.

2.1.1 The search for covalent intermediates

Experiments were designed to investigate the possibility of covalent intermediates in the mechanism, which may better account for the transpeptidation activity of the aspartic proteases than the non-covalent mechanisms. The experiments were directed at locating an acyl intermediate, as the amino intermediate has been firmly discounted on a number of grounds (see Section 1.3.1.1). Also, it had been proposed[50] that the acyl intermediate was the central intermediate in catalysis. It was decided to attempt to trap this covalent 'acyl-enzyme' intermediate (see Section 1.3.1.2), using a nucleophilic species. Experiments to trap the acyl intermediate had previously been carried out, unsuccessfully, using exogenously added nucleophiles.[39] It was decided that a more fruitful approach might be to release a highly nucleophilic species at the active site by enzymic turnover. Thus, a substrate analogue was required that would, on turnover, release a nucleophilic species at the active site.

Acylated aromatic hydrazines (mimics of substrate dipeptides) were chosen as the mechanism-activated inhibitors. If they are hydrolysed by the enzyme they would release an aromatic hydrazine at the active site. The highly reactive hydrazine may then react with the acyl intermediate as shown below (Scheme 2.1.1). A covalent complex formed between the enzyme and the hydrazine released can easily be

identified by UV spectroscopy on the inhibited enzyme. The nitrophenyl hydrazides may be preferred as inhibitors in this respect due to the very strong absorption of the nitrophenyl ring in the near UV and visible regions. This will allow straightforward detection of any protein labelled with nitrophenyl hydrazine.



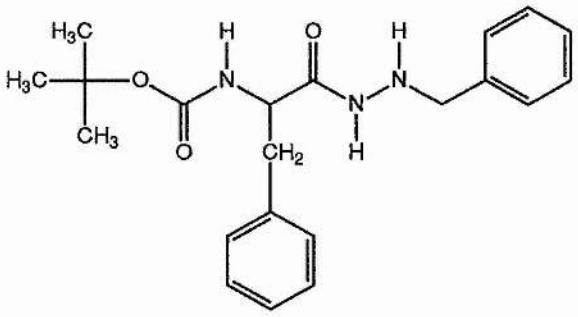
Scheme 2.1.1: Theoretical mechanism for the action of the acyl hydrazides

In line with the well-known preference for pepsin to have aromatic residues in its S_1 and S_1' sites[34] it was decided to synthesise the trapping agents with a phenylalanine residue at P_1. Originally only the phenyl and nitrophenyl versions were to have been used. However, these are 'frameshift' inhibitors (or substrates), in that the aromatic ring of the hydrazine, which corresponds to the side-chain of the carboxy terminal amino acid of a dipeptide, is one atom too close to the scissile bond. As shown below (Figure 2.1.1), the benzylhydrazide is a much better mimic of a genuine dipeptide substrate.

The hydrazine should compete efficiently with water for attack on the anhydride. It is far more nucleophilic, although the stereochemistry of attack on the acyl intermediate is probably not optimal, as it probably is for the active site water molecule. Closure of the so-called 'flap' over the active site on substrate binding (see Section 1.5.1) is thought to exclude water from most of the active site during the catalytic cycle. Thus, any nucleophile at the active site will be more effective due to the decreased solvation. Also, exclusion of water would remove a competing

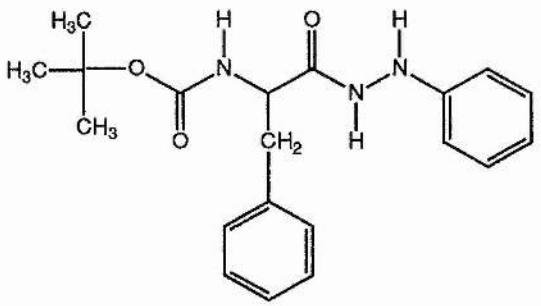
nucleophile from the active site, increasing the probability of trapping by the hydrazine.

The resulting acyl hydrazide is much more stable than the anhydride intermediate and should therefore be amenable to isolation. Thus, if it could be shown that the hydrazine is covalently bound at the active site of the enzyme, this would be very good evidence for the existence of a covalent anhydride at the active site. A similar approach to inhibiting the HIV-1 protease has since been tried and was found to fail.[268]



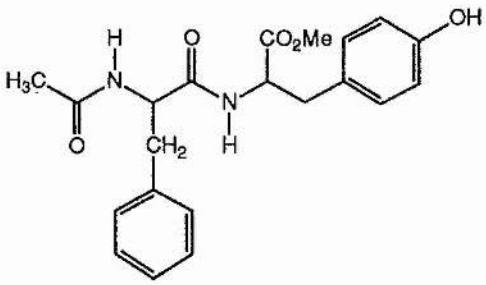
Boc-L-phenylalanyl benzyl hydrazide

Putative inhibitor



Boc-L-phenylalanyl phenyl hydrazide

Putative inhibitor



Methyl N-acetyl phenylalanyl tyrosinate

Substrate

Figure 2.1.1: The hydrazide trapping compounds

The synthesis of the *para*-nitrophenylhydrazide from *para*-nitrophenylhydrazine and methyl Boc-L-phenylalaninate presented some problems. The reaction showed little, if any, sign of progress by t.l.c. in various solvents. N.m.r. spectroscopy showed little dimunition of the methoxy peak at 3.74 p.p.m., which would have been indicative of the reaction proceeding. Various methods were attempted to force the reaction, including long reaction time (up to 72 hours reflux), higher boiling solvents

(*t*-butanol and toluene) and enhanced concentrations of reactants. None of these methods could be shown to improve the reaction. It was thus decided to explore the same reaction with two other hydrazines, namely phenylhydrazine and benzylhydrazine. Both of these should be much more nucleophilic than the *para*-nitrophenylhydrazine as the highly electron-withdrawing nitrophenyl group is in conjugation with the lone pairs of the hydrazine nitrogens. In the case of phenylhydrazine the much less electron withdrawing phenyl ring is in conjugation with the lone pairs and with benzylhydrazine the lone pairs are not conjugated at all. The reaction with hydrazine itself was also examined and could be unambiguously shown to have gone to completion by the disappearance, within 2 hours, of the methoxy peak from the proton n.m.r. spectrum of the reaction product. With phenylhydrazine the reaction was much less facile, but gave about a 50% reduction in the integral of the methoxy signal after 24 hr. reaction. Benzylhydrazine gave complete reaction overnight. Thus, the use of the nitrophenylhydrazine was abandoned in favour of the more reactive benzylhydrazine.

The reactions of phthalic anhydride and succinic anhydride with phenylhydrazine and nitrophenylhydrazine were used as models for the proposed reaction shown above (Scheme 2.1.1). These two anhydrides were chosen as models for the extensively hydrogen-bonded and highly constrained nature of the active site (see Section 1.4). The reaction of nitrophenylhydrazine with phthalic acid proceeded, as expected, very quickly. The reaction with succinic anhydride was much slower, not being complete after four days. An n.m.r spectrum of the crude reaction mixture showed that the reaction had proceeded only partially. The reactions with phenylhydrazine proceeded more quickly. The reactions with benzylhydrazine were quicker still (being complete in a few hours, depending on the solvent used).

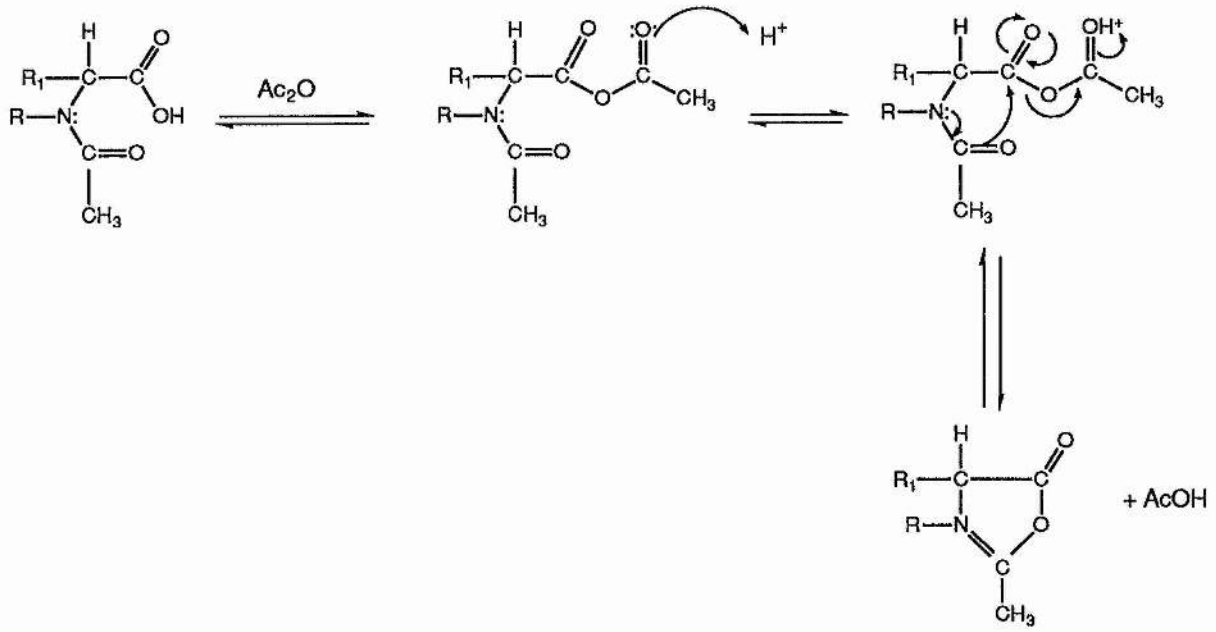
The putative acyl intermediate is probably much more like succinic anhydride than phthalic anhydride (as it retains a degree of conformational freedom that the phthalic anhydride lacks). Given the sluggish nature of the reaction between nitrophenylhydrazine and succinic anhydride it seemed that the nitrophenylhydrazine would be unlikely to react quickly enough with the acyl intermediate to give a detectable amount of trapping product. The reaction of the benzyl or phenylhydrazine may have been rapid enough to observe some trapping. Also, the enzyme active site is desolvated, which would make the hydrazines much more nucleophilic than they are in solution. This, coupled with their proximity to the acyl intermediate suggested that there was a reasonable chance to detect trapping.

The synthesis of acylated phenylhydrazines (and other mono-substituted

hydrazines) extends back to the work of Fischer,[269] who produced 1-acetyl 2-phenylhydrazines by reacting phenylhydrazine and acetic anhydride. Since that time several methods have been applied to the acylation of substituted hydrazides. Buzan *et al.*[270] used DCC and acetic acid to acetylate phenylhydrazine and Weygard and Steglich[271] used DCC (and cyanomethyl esters) to make amino acid tritylhydrazides. DCC (with HOBT) has also been used to couple hydrazine to the free carboxyl termini of peptides.[272] Yur'ev *et al.*[273] used tetrachlorosilane to activate carboxyl groups (by the formation of tetracyloxysilane) to attack by phenylhydrazine to give acylated phenylhydrazines. Kelly *et al.* [274,275] and others[276] have used the azide method and Hoffmann *et al.*[277] used a method involving 4-substituted 2-thio 5-thiazolidones. The most utilised method is an enzymatic synthesis (after Bergmann and Fraenhal-Conrat[278]) of phenylhydrazides directly from the amino acid using papain in a cysteine containing buffer. This method has been used by Waldschmidt-Leitz and Kuhn[279] and Milne and co-workers.[280] However, in common with Boinonas St. Guttmann *et al.*[281] it was decided that the mixed anhydride method would be most suitable as it is clean and efficient. The DCC mediated coupling gave purification difficulties in my hands.

It was also decided to synthesise an assay peptide for pepsin, to allow its activity to be assessed before and after the hydrazide inhibition experiments. The peptides chosen were N-acetyl-L-phenylalanyl-L-tyrosyl methyl ester or N-acetyl-L-phenylalanyl-L-tyrosine, after Clement.[47] They were chosen for their ease of synthesis and the straightforward assay method (spectrophotometrically at 237 nm).

Various methods were investigated for the synthesis of N-acetyl-L-phenylalanine. Stirring L-phenylalanine at room temperature with 1.1 equivalents of acetic anhydride in acetic acid (an initially considered method) produces extensive racemisation. This racemisation by acetic anhydride or trifluoroacetic anhydride had been reported in several cases,[282] presumably due to azlactone formation[282e,f] (see Scheme 2.1.2). Azlactone formation promotes racemisation because the alpha proton is made very acidic, due to the adjacent carbonyl and vinyl groups, and consequently is easy to remove. Reprotonation occurs from either side of the resulting double bond, giving racemisation at the alpha carbon. However, there are reported exceptions to the intermediacy of azlactones.[283]



Scheme 2.1.2: Azlactone formation in N-acetylated amino acids

It has been reported that Schotten-Baumann conditions give good yields and no racemisation,[282d,g] so this was tried (the use of pyridine and acetic anhydride was rejected as it is reported to result in at least partial racemisation[284]). There are several alternative methods that could have been deployed, for instance acylase I and sodium acetate,[285] ketene[286] or pentaflurophenyl acetate.[287] Eventually the optimal method was found to be the use of N-methoxy diacetamide,[288] as the reaction proceeds without racemisation and in good yield. A variant of the mixed anhydride method was used for the coupling.

The coupling method also had to be carefully chosen. The azide method[289] was considered as it can be used to couple unprotected amino acids. Therefore, methyl N-acetyl-L-phenyalaninate was synthesised and reacted with hydrazine to give the acyl azide. Treatment with nitrous acid at 4 °C (generated *in situ* from sodium nitrite and hydrochloric acid) gave the acyl azide, which was very unstable. Reaction of the acyl azide, without purification, with the unprotected L-tyrosine should have given the dipeptide. However, the acyl azide was difficult to isolate and the reaction proceeded in poor yield (10-20%). Thus, mixed anhydride coupling with methyl L-tyrosinate was used instead, giving an isolated yield of the dipeptide of 50%. Methyl ester protection was tried and it was found that saponification with sodium hydroxide on the model

on the model compound methyl N-acetyl-L-phenyalaninate did produce racemisation (as judged by the optical rotation of the isolated N-acetyl phenylalanine). Thus deprotection was carried out using potassium carbonate in aqueous methanol, which had been shown by model studies to produce much less racemisation. Once synthesised the peptides were tested in the assay[47] and were found to give unreliable results.

Both of the acyl hydrazides discussed (Figure 2.1.1) were incubated with pepsin under a variety of conditions. However, neither of the acyl hydrazides gave any consistently detectable loss in activity, even after prolonged incubation. Also, the addition of exogenous hydrazine (at concentrations up to 1 mM) failed to reduce activity. Furthermore, it proved impossible to unequivocally demonstrate turnover of the acylated hydrazides. Attempts to identify the products of turnover, Boc-L-phenylalanine or free aromatic hydrazine, either by difference U.V. spectroscopy or by t.l.c. using a variety of solvent systems, were unsuccessful.

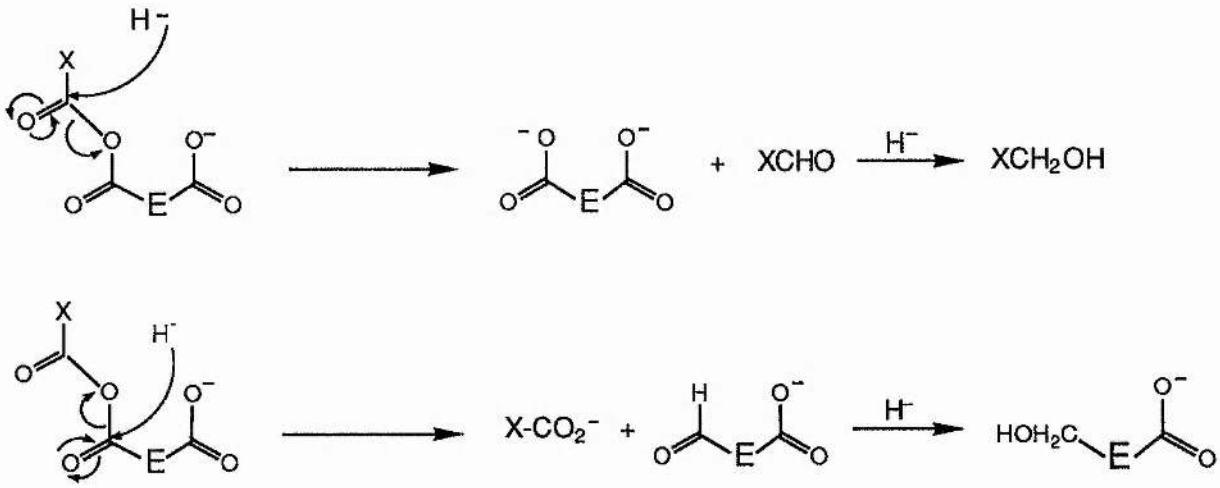
The difficulty of reliably assaying pepsin proved a handicap in these studies. Three different assays were used. The haemoglobin assay, based on Knowles'[290] modifications of the method of Anson,[291] used for much of this work was found to be difficult to reproduce. The dipeptide based assay also proved to be irreproducible. The few results obtained from assays using the nitrophenylalanine containing assay peptide (Cbz-His-Phe(NO_2)-Phe-OMe, see Experimental section) were insufficient to generate a firm model for the interaction between pepsin and the acyl hydrazides.

The reason for the lack of observable turnover is not clear. Dipeptide substrates for pepsin containing phenylalanine at P_1 usually have K_M's around 1-5 mM,[292] thus the hydrazides were used at concentrations from 10 mM downward. This should be a sufficient concentration to ensure that binding will take place. The conclusion from these studies appears to be that the hydrazides are not bound by pepsin at all.

After the failure to obtain any positive results from the mechanism generated hydrazines a different method was attempted. The work of Hofmann and Fink[59] on cryoenzymological methods for identifying covalent intermediates in aspartic protease catalysis suggested a different approach. By reducing the temperature of incubations of pepsin in methanol they hoped to slow down covalent intermediate

al.[293] made extensive use of the dipeptide L-leucyl-L-tyrosyl amide in their investigations of the mechanism of pepsin transpeptidation as it gives very high levels of transpeptidation products. If any covalent intermediate is formed in the reactions leading to cleavage (transpeptidation or hydrolysis) then this peptide should give maximum concentrations of the intermediate at pseudo-equilibrium, thus maximising the probability of trapping it.

This peptide was used as the substrate in a series of low-temperature quench experiments. In these experiments low temperatures were used not only to try and suppress the hydrolysis of the covalent intermediate but also to promote a trapping reaction with an exogenous nucleophile. Trapping was carried out by rapidly injecting an aliquot of an incubation of pepsin and the substrate into an organic solvent containing sodium borohydride as the nucleophile source. It was thought that if any active-site anhydride (see Scheme 2.1.3, where X is leucine) is formed during the course of substrate cleavage then it would be susceptible to borohydride reduction. In contrast the free acid groups of the substrate or enzyme would not be reduced. This reduction could produce either or both of two products, depending on the site of attack by the hydride anion. Reduction of the substrate derived carboxyl group would give leucinol (XCH_2OH) and reduction of the active site carboxyl would give the active site alcohol (**E**-CH_2OH).



Scheme 2.1.3: Possible reduction paths for the putative acyl intermediate
where E represents the enzyme.

The original effort was directed at identifying any leucinol formed, using t.l.c., by

comparison with an authentic sample. This approach can be modified so that detection is performed using high field n.m.r. spectroscopy, which was expected to be a more sensitive method. As a check for the occurrence of transpeptidation the initial products of transpeptidation of Leu-Tyr-NH_2, Leu-Leu, Leu-Leu-Leu and Leu-Leu-Tyr-NH_2[293] were synthesised and used as standards for t.l.c on silica plates. The developing system used was that of Antonov,[294] MeOH: EtOAc: NH_3 (35:65:3).

The peptides were synthesised using Boc and methyl ester protection and coupled *via* the mixed anhydride method.[295] The use of DCC and the azide method had been explored previously and were found to be much inferior to the mixed anhydride method. This strategy was chosen over Cbz protection as Cbz-L-leucine is an oil and is much more difficult to isolate pure than Boc-L-leucine.

The experiments were performed at a variety of temperatures (-10 to -25 °C), in a range of solvent mixtures (0-20% ethanol in THF) containing sodium borohydride (0.5 to 5 mM). The lowest possible temperature was necessary to try and slow the rate of hydrolysis of the acyl intermediate on denaturation of the protein in the organic mixture. Incubation of the Leu-Tyr substrate and pepsin at varying concentrations was carried out and aliquots removed at intervals and added to a large excess (25 fold) of the cryo-quenching mixture. Much care was taken to ensure that the aliquot did not freeze on contact with the cold organic mixture, as then the slow thawing would result in complete hydrolysis of the acyl intermediate before it could be trapped by borohydride. Temperatures lower than -25 °C always resulted in a degree of solidification of the aliquot on contact so most experiments were carried out at -25 °C. Control samples consisting of only Leu-Tyr amide or only pepsin were also run for comparison.

After low temperature trapping, the organic solvents were removed under reduced pressure and a portion of the remaining aqueous phase applied to a t.l.c. plate and developed as above. The remainder was desalted on a small Sephadex G-25 column and the protein containing fractions (which absorb at 280 nm) were collected, lyophilised and analysed as discussed below. Unfortunately, although some of the transpeptidation products could be identified by t.l.c. no leucinol could be unambiguously identified. Organic extracts of the aqueous phase were also taken, dried, evaporated and analysed by n.m.r. and further t.l.c., and still no leucinol could be observed. The possibility of trapping of the anhydride by ethanol in the cryo-quenching solution, to give leucine ethyl ester or an esterified active site

carboxyl, was also considered. Quenching in neat THF avoided this problem.

Attention was turned to the protein, and attempts were made to identify an 'altered' form of pepsin by comparisons of mobility on denaturing polyacrylamide gels.[296] Isoelectric focussing would probably have proved the more useful technique as the reduction of an active site carboxyl group will alter the overall charge on pepsin. This charge will be reduced still further if lactonisation occurs with the remaining active site carboxyl group.

To facilitate the observation of any pepsin of altered mobility it was decided to develop a method to remove the phosphate group on serine-68. This gives rise to two bands on SDS gels, one due to the phosphorylated protein and one to the non-phosphorylated form. The faster running form is probably dephosphorylated.[297] Perlmann removed the phosphate group from pepsin and pepsinogen using intestinal alkaline phosphatase.[298] Initial dephosphorylation experiments on desalted pepsin from the quench experiments and on native pepsin were carried out using commercial alkaline phosphatase. However, this gave poor results so the phosphatase was further purified by DEAE-cellulose chromatography, followed by ultrafiltration. The concentrate was then used for further experiments, which still failed to give one band from pepsin on the SDS gels.

A consistent problem with this approach was the difficulty in visualising the pepsin run on SDS gels. Pepsin was poorly and irregularly stained with Coomassie Brilliant Blue G250, as has been noted by several authors.[299] This effect may be due to the low level of basic amino acids in pepsin (1 lysine, 2 arginine and 2 histidine from 327 in total[300]). Similarly, zeins, proteins isolated from corn seed that contain almost no lysine and little (around 3%) arginine, have been shown[301] to give very low absorbance readings in the Bradford assay.[302] The binding of Coomassie Blue to proteins has been shown to correlate with both the basic and hydrophobic amino acid content,[303] so poor binding may not just be due to low levels of basic amino acids.

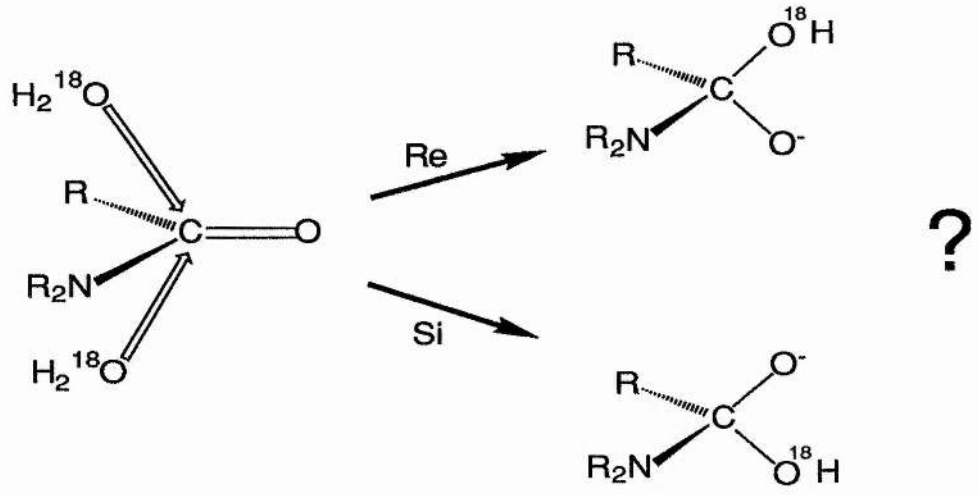
Other stains tried included Naphthol Blue Black and Fast Green FCF. Fast Green proved to be the best of those tried, however, the staining was still not very sensitive or reliable. The irregular staining problem could also be partially overcome by loading the gels at much higher levels than is normal (protein concentrations of 4-16 mg ml^{-1} were used compared with the normal range of 0.2-2 mg ml^{-1}). None of these experiments could be shown to provide evidence for formation of pepsin with altered

experiments could be shown to provide evidence for formation of pepsin with altered electrophoretic behaviour. It is expected that the loss of one carboxyl group from pepsin (as it is converted to the alcohol) would result in a change in its electrophoretic mobility. However, given that there are 43 acidic residues in pepsin,[300] the loss of just one may well have no observable effect.

2.1.2 Face of attack of the lytic water molecule

A still unresolved question in the nucleophilic water mechanisms is the orientation of the scissile carbonyl within the active site. Does it point in toward the active site aspartates, or outward toward the flap? Several theories favour the first alternative,[21,78,81] others the latter.[80] In an attempt to resolve this conflict an experiment was designed to attempt to trap the tetrahedral intermediate proposed in these mechanisms. At the same time it was hoped that stereochemical information on the direction of water attack on the scissile carbonyl could be derived.

The two possible modes of attack by the lytic water molecule are shown below (Scheme 2.1.4).



Scheme 2.1.4: Two possibilities for the direction of water attack upon the scissile bond

The two possible directions for attack upon the scissile carbonyl by ^{18}O labelled water result in two possible isotopomers of the tetrahedral intermediate, as shown

above. It was hoped that this intermediate could be trapped by positioning a group close to it that possessed a suitably oriented leaving group. Thus, a substrate for this experiment would be a peptide with an amino acid at P_1' that had a side-chain modified to contain a good leaving group. The isosteric valine analogues, the bromobutryines,[304] lent themselves ideally to the task.

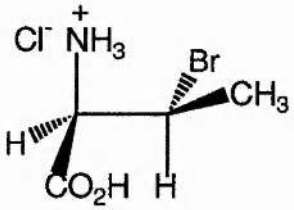


Figure 2.1.3: (2S, 3S) 2-amino-3-bromobutyrine hydrochloride

This compound and its C-3 epimer, the 3R compound, have bromine as a good leaving group that it was hoped would be suitable for trapping the tetrahedral intermediate. An attempt was made to predict which of the two epimers of the bromobutyrine would be the most likely to trap the tetrahedral intermediate using molecular modelling. A crystal structure of unliganded pepsin[126b] with a tripeptide, N-acetyl-L-phenylalanyl-bromobutyrine-L-alanine (Figure 2.1.4) manually docked into the active site was examined. No good guide as to the most favourable epimer for trapping could be obtained.

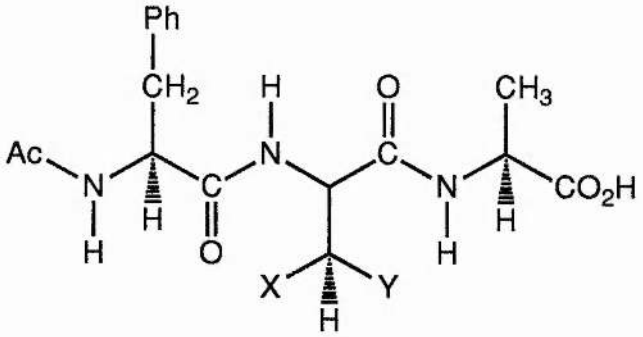


Figure 2.1.4: The trapping peptide

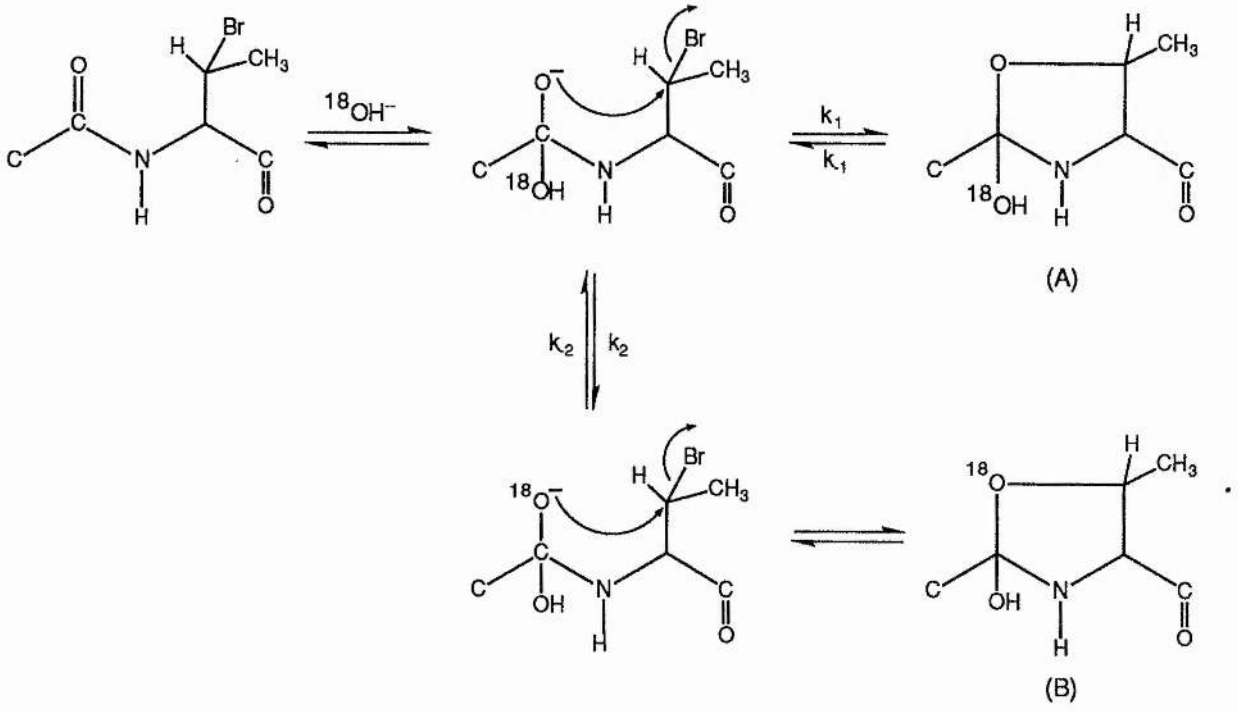
X = Br, Y = CH_3 - 3R epimer

X = CH_3, Y = Br - 3S epimer

This sequence for the trapping peptide was chosen as phenylalanine is known to

that these favoured residues would ensure that the peptide bound productively in the P_1 and P_2' pockets and not non-productively in any other of the enzyme subsites. This will ensure that the scissile bond lies on the N-terminal side of the bromobutyrine to maximise the probability of trapping the oxyanion.

An outline of the trapping experiment is given below (Scheme 2.1.5). The experiment was to have been carried out in ^{18}O-labelled H_2O. The amount of trapping was expected to be very small indeed, so very sensitive methods would have been needed for detection. Mass spectrometry seemed ideally suited to this task as only small amounts of material would be needed and the sensitivity of detection of any ^{18}O-label incorporated into the product lactims would be high.



Scheme 2.1.5: Trapping of oxyanion intermediate by bromobutyrines

The products will be isolated as the lactims (C) (from product (A) above) and (D) (from product (B) above) and analysed by mass spectrometry.

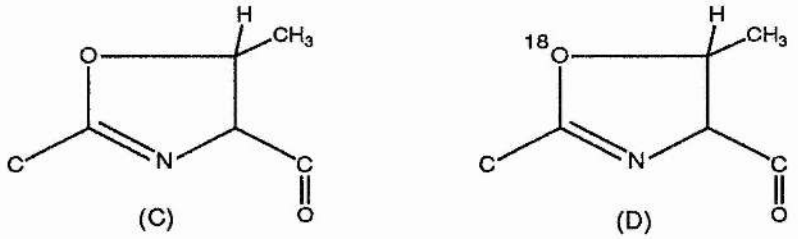
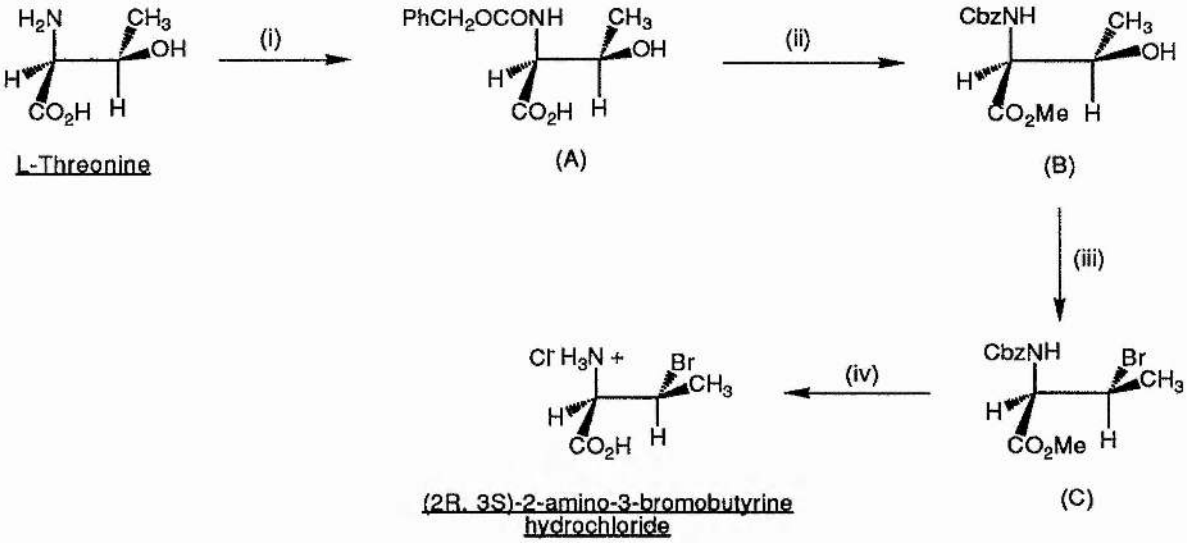


Figure 2.1.5: Lactims, the proposed final product of the bromobutryine oxyanion capture experiments

It is assumed that the two oxygen atoms of the *gem*-diolate are in rapid protonic equilibrium, at a rate much faster than either k_1 or k_2 in Scheme 2.1.5. Which of the two epimers of the bromobutyrine gives the most incorporation of ^{18}O-label into the lactims will provide information on the face of approach of the lytic water molecule to the scissile carbonyl bond.

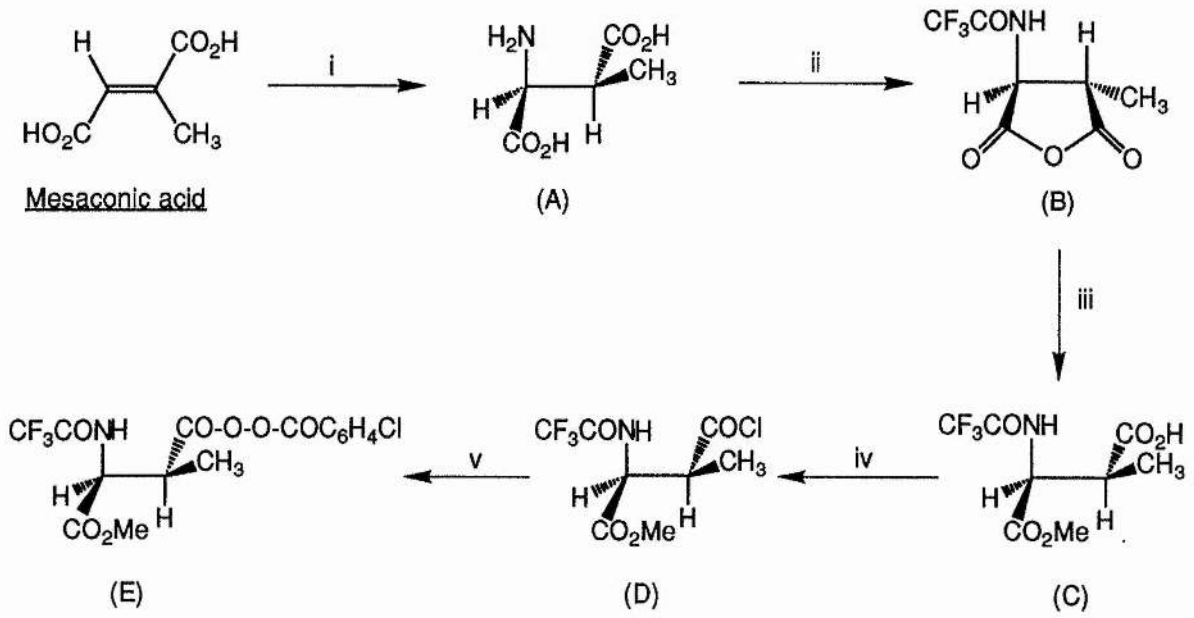
The syntheses of the two epimers of the bromobutyrine were carried out initially following the method of Akhtar and Gani,[304] which is outlined in Scheme 2.1.6.



Reagents: (i) $PhCH_2OCOCl$, $NaHCO_3$, water, 25°C, 3 h.
(ii) CH_2N_2, diethyl ether, 25°C, 1 h.
(iii) CBr_4, PPh_3, benzene, 25°C, 90 min.
(iv) 12 M hydrochloric acid, glacial acetic acid (1:1 v/v), 100°C, 2 h.

Scheme 2.1.6: Synthesis of the (2R,3S) bromobutyrine

If the same approach was to be followed for the synthesis of the 3R epimer then a source of L-*allo*-threonine, the starting material, was required. This is available commercially but is prohibitively expensive, so a synthetic route was required. The initial approach is shown in Scheme 2.1.7.



Reagents: (i) β-Methylaspartase, Mg^{2+}, K^+, H_2O, 30°C, 45 h.
(ii) $(CF_3CO)_2CO$, THF, 25°C, 2 h.
(iii) Methanol, 25°C, 20 mins.
(iv) $SOCl_2$, 80°C, 1 h.
(v) 85% *m*CPBA, pyridine, Et_2O, 0°C, 4 h.

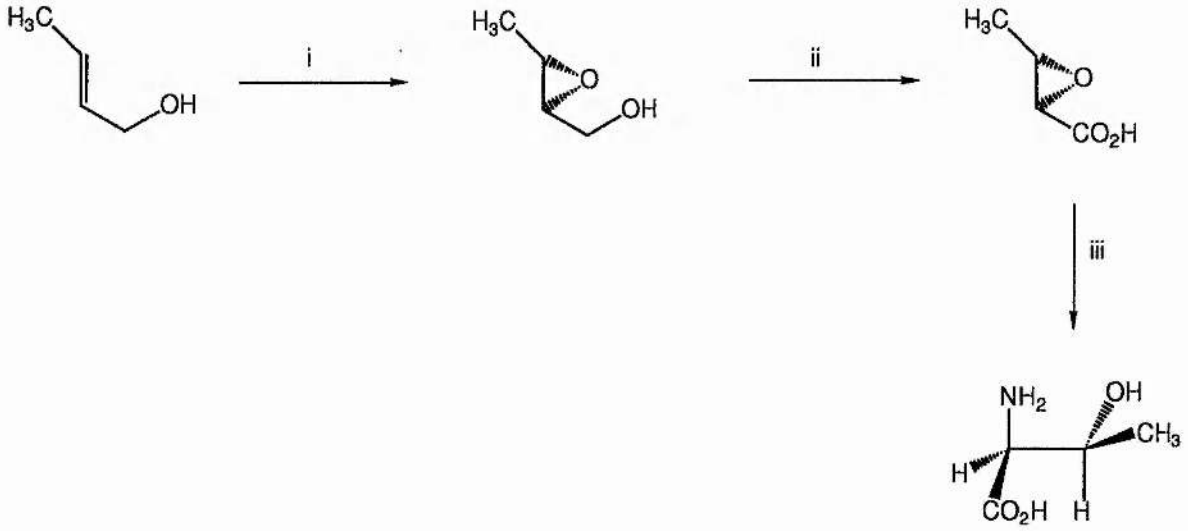
Scheme 2.1.7: The attempted synthesis of L-allo-threonine

The final two steps of the synthesis, which were not achieved, are the thermally promoted radical rearrangement of the perester to give N-protected, O-*m*-chlorobenzoyl methyl L-*allo*-threoninate, which would then be subjected to acid hydrolysis, to give L-*allo*-threonine hydrochloride.

Compound (C) is produced as a mixture of the α ester and the β acid and the α acid and the β ester as shown by 2 methoxy peaks in the proton n.m.r. spectrum. The two regioisomers may be resolved by a fractional crystallisation, which proved to be extremely difficult. In an attempt to increase the selectivity of the opening of the

methylaspartic acid anhydride the reaction was carried out at 0 °C and -10 °C, but without significant effect. The mixture was carried through to reaction with the *m*CPBA, but the mixture of regioisomers from this reaction could not be separated either. Once the L-*allo*-threonine is obtained the synthesis of the (2R, 3R) bromobutyrine would proceed exactly as for the (2R, 3S) diastereomer.

The most obvious route to L-*allo*-threonine is a Mitsunobu-type inversion of the C-3 centre of L-threonine, however this has been shown to result in elimination.[306] Thus an alternative synthesis of the starting material was required. It was considered that the approach outlined in Scheme 2.1.8 might provide L-*allo*-threonine cheaply in reasonable quantity. The initial step is difficult, as the usual aqueous work-up of Sharpless epoxidations cannot be used due to the high water solubility of the epoxide product.[307] Sharpless oxidation[308] and opening of the epoxide with ammonia would then complete the synthesis.



Reagents: (i) tBuOOH, $Ti(iPrO)_4$, (2R, 3R)-(+)-DIPT, CH_2Cl_2, -78 °C.
(ii) $RuCl_3.3H_2O$, $NaIO_4$.
(iii) NH_3.

Scheme 2.1.8: The new synthesis of L-allo-threonine

However, whilst this was in progress the same route was published.[309]

In the initial synthesis of the bromobutyrines (Scheme 2.1.6) the final step is acid

hydrolysis of the amino and carboxy protecting groups in one step, yielding the bromobutyrine hydrochloride. This is unsuitable for further coupling as it is carboxy unprotected. This acid deprotection step is necessary as the bromine-carbon bond is not stable to the hydrogenation conditions usually employed to remove the benzyloxycarbonyl group. As a result the bromobutyrine methyl ester, which is required for further peptide coupling, cannot be produced directly from the fully protected bromobutyrine if benzyloxycarbonyl amine protection is used. The carboxy group could be reprotected after the acid deprotection, but this is inefficient.

Unprotected amino acids can be coupled in weakly alkaline aqueous solution by the azide coupling method. This method was explored using the model coupling of acetyl-L-phenylalanine and L-valine. The azide method has the additional attraction that it has been reported that it is the only method to provide racemisation-free coupling of N-acetylated amino acids.[310] However, poor yields resulted and the procedure did not appear well suited to the model system. The carboxy protection chosen for the L-valine residue had to be carefully chosen as no base-sensitive groups could be used or the base employed may have racemised the N-acetyl-L-phenylalanine. Thus, acid-sensitive *tert*-butyl protection was investigated. Several methods were investigated for synthesising *tert*-butyl-L-valinate. These included acid-catalysed esterification with liquid isobutylene[311] and reaction of the acid chloride or mixed carboxylic anhydride[312] with *tert*-butanol and/or sodium *tert*-butoxide. None gave significant amounts of product (as judged by the lack of a *t*-butyl signal around 1.5 p.p.m. in the proton n.m.r spectrum) so the approach was abandoned. The use of *tert*-butyl protection would have allowed selective removal of a base-sensitive N-protecting group. Consequently it was decided that Fmoc[313] (9-fluorenylmethyloxycarbonyl) and or Boc (tertiary butoxycarbonyl) amino protection would be used, which can be easily removed without affecting the carbon-bromine bond.

Fmoc seemed ideal as it is easily put on and removed and its base, rather than acid, lability means that the free amine product is produced, which can be used directly in the coupling procedure. It also allows the carboxy blocking group, in this case a methyl ester, to be retained right through the synthesis, thus increasing the overall yield. The Fmoc group was put on using Fmoc chloroformate[313] and no production of dipeptide was observed thus eliminating the need to use Fmoc azide.[314] Similar advantages arise from the use of Boc protection, except that the

amine salt arising from the acid catalysed removal of the Boc group would need to be neutralised prior to coupling. Both approaches were tried and the Boc group proved superior due to greater ease of handling of some of the synthetic intermediates. Unfortunately, the synthesis of the peptides for the experiment was not completed as attention was directed elsewhere.

In summary the series of experiments carried out on pepsin were inconclusive or incomplete. The attempts at trapping a possible covalent intermediate gave only negative information. After the trapping experiments were carried out the aspartic protease from HIV was brought to our attention. It seemed an attractive target for a variety of mechanistic and active site directed studies due to its small size (see Section 1.6.2) and availability. It was hoped that information gleaned from these studies would be of use in the design of therapies for HIV infection, based on inhibition of the protease.

2.2 Design of HIV-1 Protease Inhibitors

In undertaking the design of a novel class of HIV protease inhibitors the strategy developed so successfully in the design of renin inhibitors was followed.[226] In this strategy a good peptide substrate sequence is identified and its scissile amide bond replaced by a non-cleavable analogue. A good substrate sequence can be identified from a consensus of sequences of preferred peptide and/or protein substrates. It is assumed that good substrate sequences will provide strong binding for the inhibitor as all the binding energy available will be expressed in binding and not in distorting the ligand as occurs with substrates.

We searched for a scissile bond analogue that possessed maximum congruency to the transition state, or stable intermediate, for the reaction catalysed by the protease. A moiety was required that possessed the most similar geometry, charge, hydrogen bonding capacity and size possible. The principle underlying the insistence on close mimicry was two-fold. First, transition state analogues were expected to be bound tightly by their target enzyme.[315] Second, all other classes of scissile bond analogues so far developed lack one or more hydrogen bond donors or acceptors compared to the stable intermediate and this may have an influence on the active site structure (see below). The phosphonamidate moiety was eventually

chosen as the scissile bond mimic. When this work was begun little was known about the properties of phosphonamidates as aspartic protease inhibitors, and no phosphonamidate-based HIV-1 protease inhibitors had been published. The use of a phosphorus containing transition state analogue has the additional advantage that ^{31}P n.m.r. spectroscopy may be used to investigate rates and equilibria in binding to the protease.

As discussed in Section 1.3.3 the exact nature of the stable intermediate in amide hydrolysis by the aspartic proteases is unknown. A phosphonamidate and its methyl ester with the two possible tetrahedral stable intermediates are shown in Figure 2.2.1. It can be seen that they have very similar charge distributions, geometry and size. Phosphorus-oxygen single and double bonds are about 10-15% longer than the corresponding carbon-oxygen bonds.[316] This allows the P=O double bond in the phosphonamidate to mimic the lengthening of the carbonyl bond as it is attacked by the lytic water molecule.

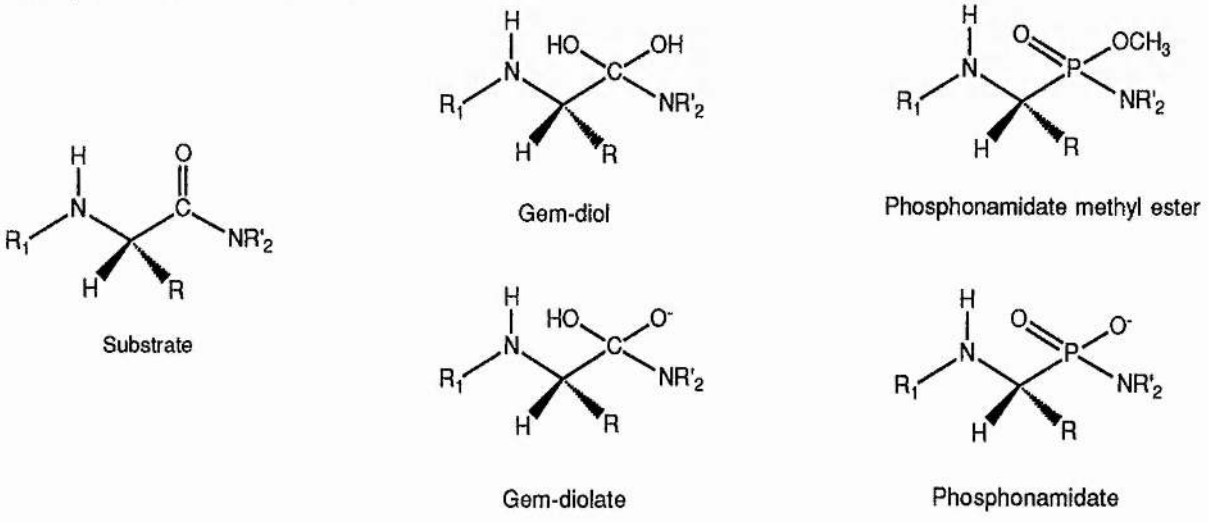


Figure 2.2.1: Comparison of the possible stable intermediates and phosphonamidates

As the pK_a for the phosphonamidate is around 3,[317] at the pH optimum for the HIV-1 protease (5.5, see Section 1.6.2) it will be negatively charged. Thus, the phosphonamidate will not be a good mimic for the neutral tetrahedral intermediate under the conditions of testing. If it should be that the stable intermediate in hydrolysis is the uncharged *gem*-diol-like amide hydrate (see Section 1.3.3) then the

phosphonamidate methyl ester would be expected to be the closest mimic. The charge congruency of the phosphonamidate methyl ester and the *gem*-diol-like tetrahedral intermediate has been proved by semi-empirical quantum mechanical calculations carried out using MOPAC 6.0 with the AM1 method.[318] The results are given below in diagrammatic form, with the partial charges residing on each atom shown. A similar calculations have been carried out[319] and similar results were obtained.

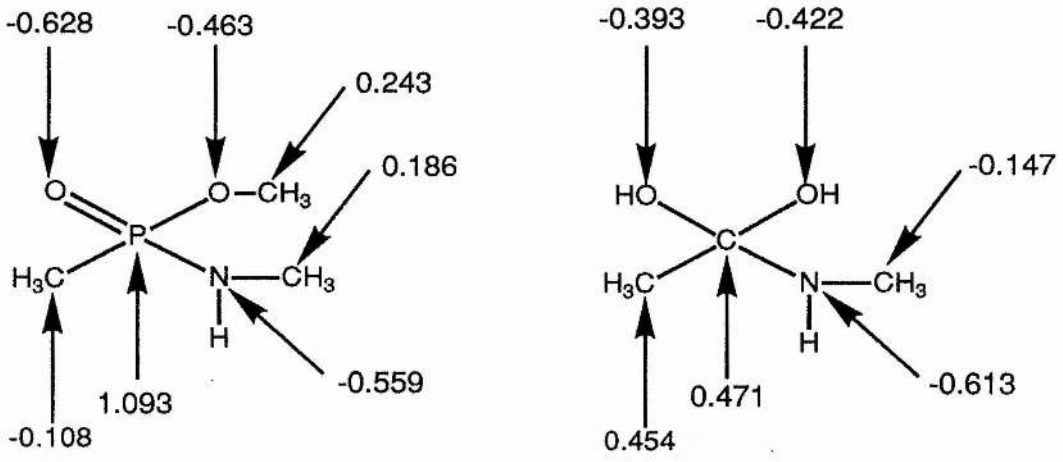


Figure 2.2.2: Charge density comparison of a model phosphonamidate methyl ester and a model amide hydrate

The crucial points of similarity are the oxygens and the nitrogen of the hydrate and of the phosphonamidate methyl ester. The partial charges on the oxygens can be seen to be similar, showing again the close analogy between the two moieties. An *ab initio* calculation was recently carried out on phosphonamidates and similar results to those given above were found.[31] In the same study it was found that the stable intermediate for acid-catalysed peptide bond hydrolysis is likely to be the neutral amide hydrate (see Section 1.3.3). Phosphonamidates were not good models for this stable intermediate but phosphonamidate methyl esters were found to be better. The transition state for amide hydrolysis was found to be zwitterionic.

Only the phosphonamidate, hydroxyethylamine, norstatine and reduced peptide scissile bond mimics (see section 1.6.3) possess a nitrogen atom in the P_1' residue. The dihydroxyethylene inhibitors also have a hydrogen bond acceptor at this position (the C-terminal C-OH group). It has been proposed that the P_1' nitrogen in

HIV-1 protease substrates is involved in hydrogen bonding to the OH of aspartate 25' at the active site.[215] Possession of a hydrogen bond acceptor at this position in an inhibitor could potentially be important for maximising binding and will affect the conformation of the active site aspartates. Therefore, we considered it important to retain this interaction in our inhibitors to provide as close a mimic for the tetrahedral intermediate as possible.

The other aspect of the design to be considered was the nature of the amino acid sequence flanking the scissile bond. At the time this work commenced, most of the data available on the subsite specificity of the HIV-1 protease had been derived from studies on the *gag-pol* polyprotein cleavage sites. Thus, our inhibitors were initially based on sequences derived from these sites. A striking feature of these sites in the viral polyprotein is the number with proline at P_1' (see Section 1.6.2). This cleavage is unique to the retroviral proteases as Xaa-Pro bonds are not cleaved by the mammalian or fungal aspartic proteases.[204] This property can therefore be used to build selectivity into our inhibitors as if a Phe-Pro scissile bond mimic is used the resulting inhibitor will not be tightly bound by host proteases, only the HIV protease. Retroviral aspartic proteases all seem to require minimally hexameric substrates (see Section 1.6.2), which indicated that an inhibitor with 6 binding determinants would be required. The amino acid at P_1 was found to be nearly always large and hydrophobic.[320] Considering all the above a candidate sequence was designed:

Phe-Asn-Phe * Pro-Ile-Val

All the residues of this sequence occurred at least twice in the same position in the cleavage sites found in the *gag-pol* polyprotein, thus this sequence should be a good substrate. Replacement of the scissile Phe-Pro bond should then provide a good inhibitor. Replacing phenylalanine at P_3 by the Boc protecting group gave the sequence of the target compound 000 (Figure 2.2.3). Thus the initial target has 6 binding determinants (like a hexapeptide) but only 3 peptide bonds. A very similar design strategy has recently been used in the development of *allo*phenylnorstatine-based inhibitors for the HIV-1 protease[236] which were found to be quite potent (K_i's of the best inhibitors around 10 nM).

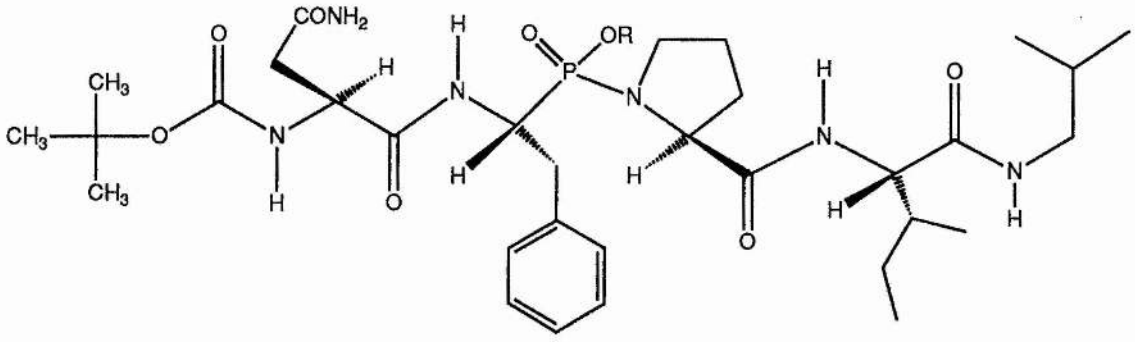
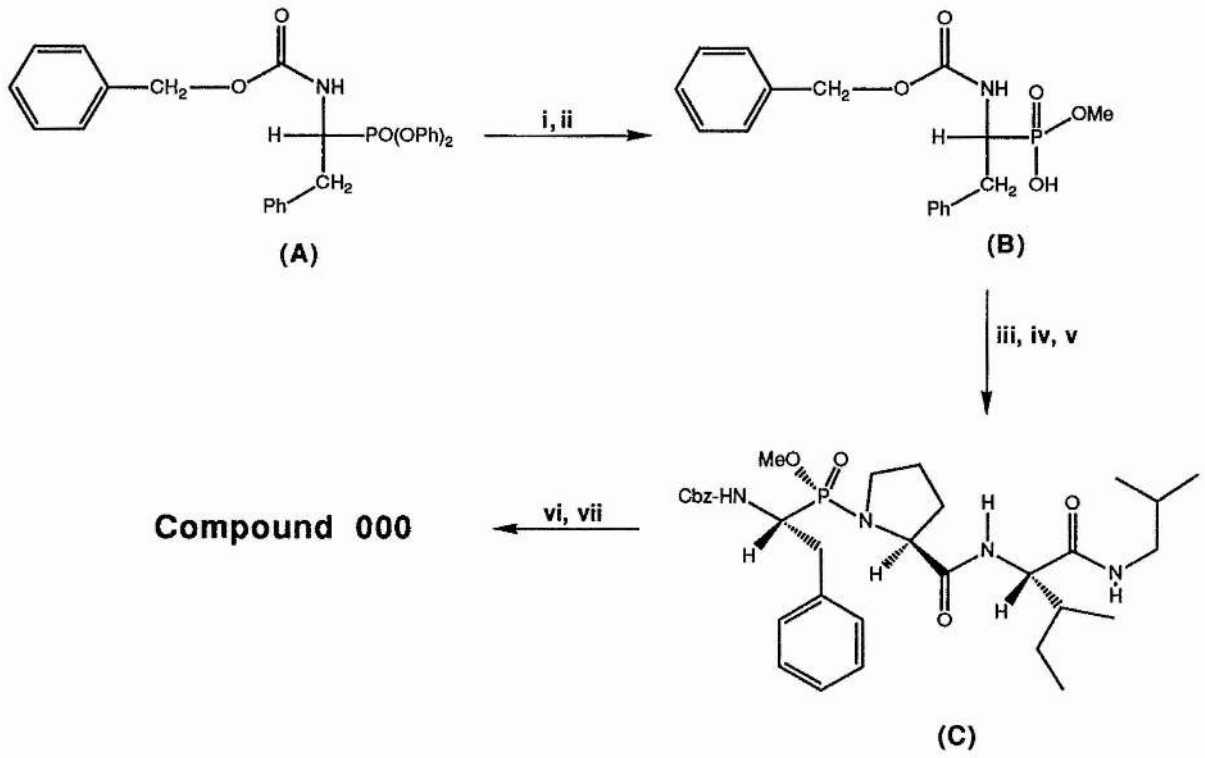


Figure 2.2.3: The initial target compound, DG 000

where R = H or CH_3

The program of synthesis, testing, modelling of the target compounds and evaluation of results was carried out in conjunction with Mr. Nicholas Camp and Mr. David Perry. The synthesis of the central aminophosphonic acid moiety and most of the *in vitro* testing was carried out by me. Some of the later assays were carried out by Mr. David Perry. The aminophosphonic acid moiety was incorporated into the inhibitors by Mr. Nicholas Camp, as shown in Scheme 2.2.1. The compounds were also tested in an *in vivo* assay by Dr. Derek Kinchington at St. Bartholomew's Hospital.



Scheme 2.2.1: The synthesis of compound 000

Reagents: (i) 2.2 eq. NaOMe, MeOH, r.t.; (ii) 2.5 eq. NaOH, MeOH, r.t.; (iii) $SOCl_2$, CH_2Cl_2, r.t. then; (iv) HN-Pro-Ile-NHiBu.HCl, NEt_3, CH_2Cl_2, r.t.; (v) Chromatographic resolution on silica EtOH/ CH_2Cl_2 5:95; (vi) Pd/C, H_2, MeOH, r.t.; (vii) Boc-Asn, iBuOCOCl, NMM, THF, -15°C.

The phosphonamidate is produced from the methyl ester (compound 000) by treatment with 4 eq. LiOH in dioxane. The synthesis of compounds (A) and (B) will be discussed in detail in Section 2.3. The stereochemistry of the 2 resolved diastereomers of compound 000 and the 4 diastereomers of the methyl ester (the phosphorus of the methyl ester is chiral) was assigned by Nicholas Camp using a combination of X-ray crystallography on a synthetic intermediate, chemical degradation, ^{1}H, ^{13}C and ^{31}P n.m.r. spectroscopy.

2.3 Synthesis of the phosphonamidate portion

The phosphonamidates were synthesised by coupling activated aminophosphonic acid half esters to various peptides, as shown in Scheme 2.2.1 above. The half

esters were activated for the coupling by conversion to the phosphoryl chloride with thionyl chloride. Therefore, an efficient route to N-protected aminophosphonic acid half esters was required. These species were then used by Nicholas Camp to assemble the inhibitors prior to testing.

2.3.1 Synthesis of aminophosphonic acids

The synthesis of aminophosphonic acids, which are close analogues of the amino acids, has attracted much interest and a wide variety of methods have been developed. The study of phosphorus analogues of the naturally occurring amino acids was begun by Chavane[321] and the first phosphonamidate containing peptides were synthesised by Imoto[322] and Martell.[323,324] Much work since then has been directed at the synthesis of 1-aminoalkylphosphonic acids and their esters and peptides containing phosphonamidate moieties (see Figure 2.3.1).

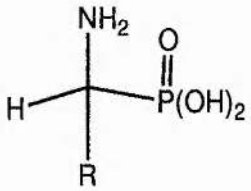
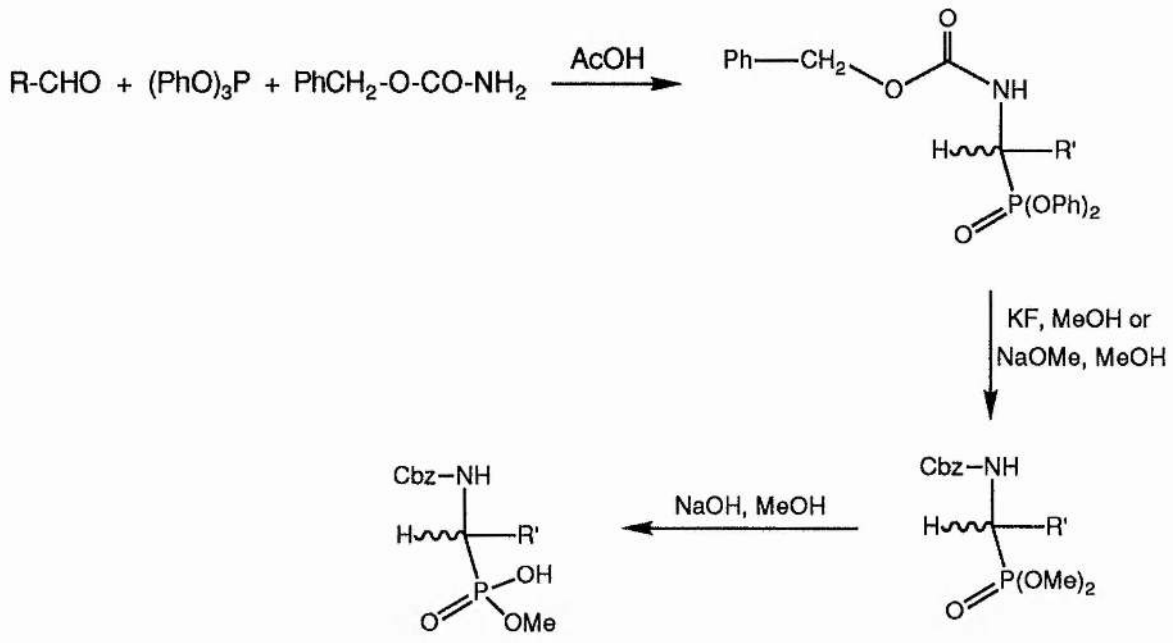


Figure 2.3.1: A 1-aminoalkylphosphonate

Also, due to their closer resemblance to carboxylic acids, the alkylphosphonous acids (which have 1 OH group) have been extensively studied and most of the analogues of the naturally occurring amino acids have been synthesised.[325]

Various approaches to the synthesis of racemic 1-aminophosphonic acids have been taken. The classical route is the addition of trialkyl phosphites to imines in refluxing toluene.[326] A milder method involves the addition of trimethylsilyl esters of phosphorus acids to imines in dichloromethane at 0 °C.[327] Recently the three-component condensation of a carbamate, aldehyde and phosphorus triester has been extensively exploited to give N-protected alkyl and aryl[328] phosphonic diesters. Also, oxidative decarboxylation of α-aminocarboxylic acids to give O-acetyl N,O acetals and subsequent reaction with a trialkyl phosphite has been used. The decarboxylation has been carried out electrochemically[329,330] or with lead

tetraacetate.[331] After considering all these alternative routes it was decided to opt for the three-component coupling developed by Oleksyszyn *et al.*[332] which gives diphenyl phosphonates. Two further reactions would then provide the required phosphonate half-esters (Scheme 2.3.1).

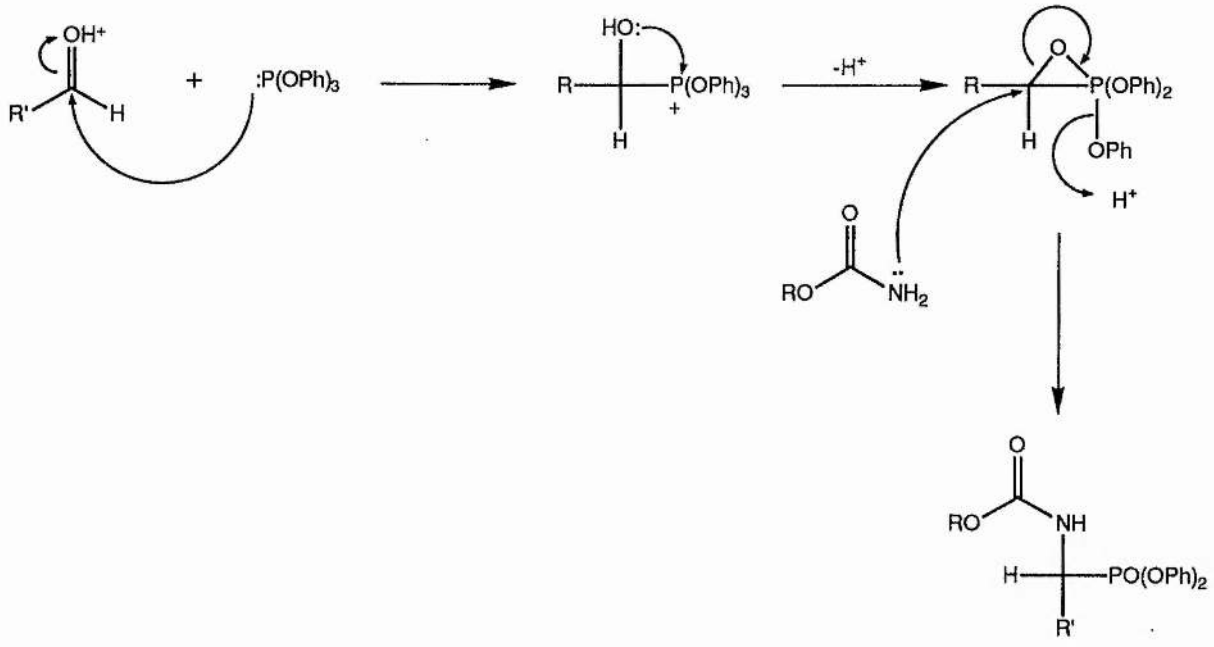


Scheme 2.3.1: Synthesis of the N-protected 1-aminoalkylphosphonate half esters

The diphenyl phosphonate was formed, in only moderate to poor yields (29-53% depending on R), by the three component condensation, then transesterified by treatment with sodium methoxide[333] or potassium fluoride in methanol[334] to give the dimethyl phosphonate. Removal of one methoxy group by treatment with sodium hydroxide and subsequent chlorination provides the phosphoryl chloride derivative, suitable for coupling to the C-terminal peptidic moiety. However, this route gave a low overall yield (around 15-25%), due to the inefficient first step.

A more obviously direct route is to use trimethyl phosphite, instead of triphenyl phosphite, in the three component coupling to give the dimethyl phosphonate directly. The removal of step two would increase the overall yield. However, when this was tried under the conditions of the three component coupling a product without methoxy group signals in the proton n.m.r. was formed. The reasons for this are unclear as some kind of reaction does definitely occur (the reaction is highly exothermic in the first few minutes) but the desired product is isolated in very low

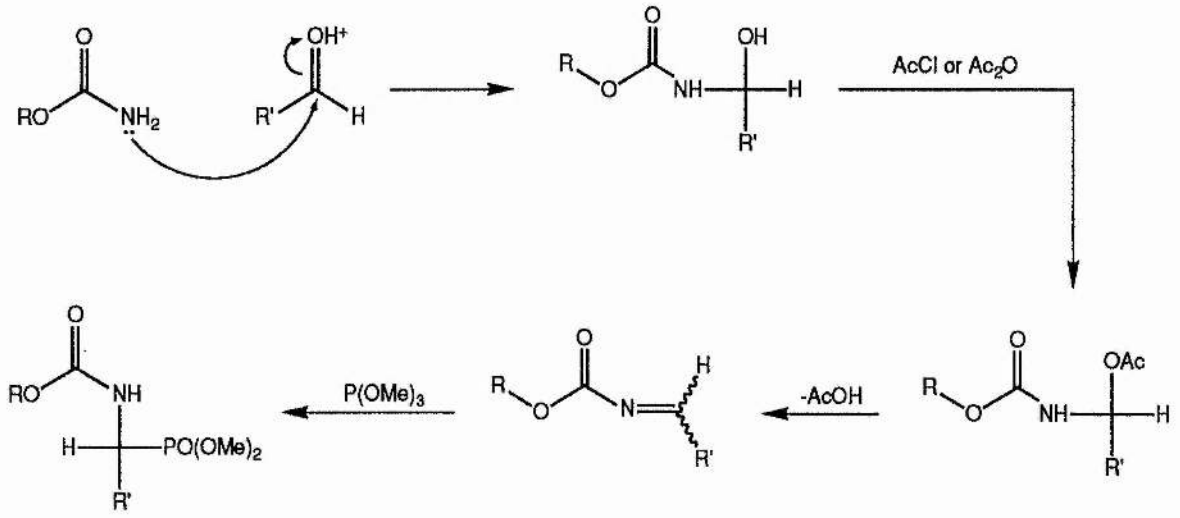
yield. As judged by ^{31}P n.m.r. the product from this reaction also contained no phosphorus. The reaction therefore does not proceed as for triphenyl phosphite, where the first step is probably attack on the aldehyde by the phosphite (Scheme 2.3.2).



Scheme 2.3.2: Possible mechanism for the three component coupling

A possible alternative mechanism involves initial attack by the carbamate (instead of the phosphite) on the aldehyde, followed by reaction of this intermediate with the phosphite (Scheme 2.3.3). This has precedent in the three component coupling involving formaldehyde.[335] Thus efforts were made to preform the possible reaction intermediate between the aldehyde and the carbamate and then react this species with trimethyl phosphite in a separate step. Since trialkyl phosphites add to imines,[326] attempts were made to synthesise the imine by acid-catalysed dehydration. Other attempts used acetic anhydride or acetyl chloride to generate the O-acetyl intermediate that would then eliminate to give the imine. These all proved fruitless, as no addition product could be detected by t.l.c or proton n.m.r. spectroscopy. A variant on this reaction was tried, using lead tetraacetate to give the imine from racemic N-protected amino acids.[336] The imine was then treated with

trimethyl phosphite or the dimethyl phosphite anion but the reaction was found to be unreliable and was therefore replaced by the three component coupling (Scheme 2.3.1).



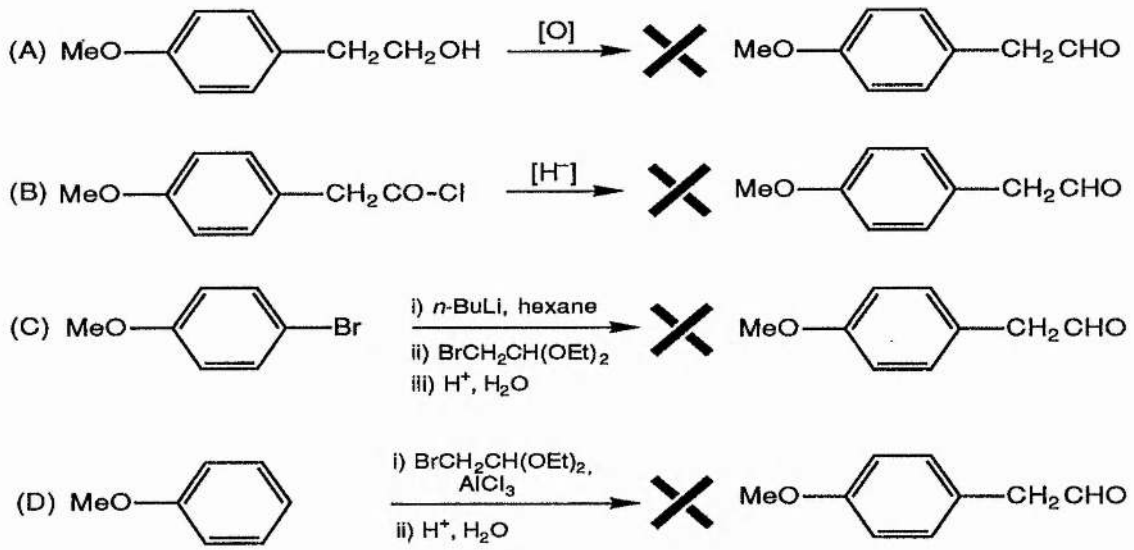
Scheme 2.3.3: Proposed reaction to give three component coupling with trimethyl phosphite

An obvious variant on the initial inhibitor sequence would be to replace the phosphophenylalanine residue at P_1 with the phospho analogue of tyrosine. This requires a protected version of *para*-hydroxy phenylacetaldehyde for the three component coupling, probably *para*-methoxy phenylacetaldehyde. This is not commercially available, so a number of possible routes to it were examined. As shown in Scheme 2.3.4 there are a number of alternatives. The most obvious is oxidation of *para*-methoxy phenethanol (route A). Several mild methods of oxidation were explored, such as the Swern,[337] Parikh[338] and Collins[339] oxidations and the use of PCC.[340] None proved to be effective, giving mixtures of products that could not be identified, as well as a small amount of aldehyde (as judged by a peak around 9 p.p.m. in the proton n.m.r.). Swern[337] found that the oxidation of 2-phenylethanol gave only a 23% isolated yield of the corresponding aldehyde, consistent with the results found in this study.

Several routes from *para*-methoxy phenylacetic acid or derivatives were considered. Reduction of the acid chloride (route B) with lithium tri-*tert*-butoxy

aluminium hydride[341] or bis(triphenylphosphine)tetrahydroboratocopper[342] proved ineffective. Reduction of the methyl ester[343] or the Weinreb amide[344] with diisobutyl aluminium hydride were also considered but not attempted.

A rather different approach (routes C and D) was also tried, joining a suitably protected aldehyde fragment (bromoacetaldehyde diethyl acetal) to the *para* position of anisole and then deprotecting. This can be done using a Grignard reagent or lithium salt generated from a *p*-haloanisole (route C), or under Friedel-Crafts conditions (route D). Neither of these routes proved to be successful.



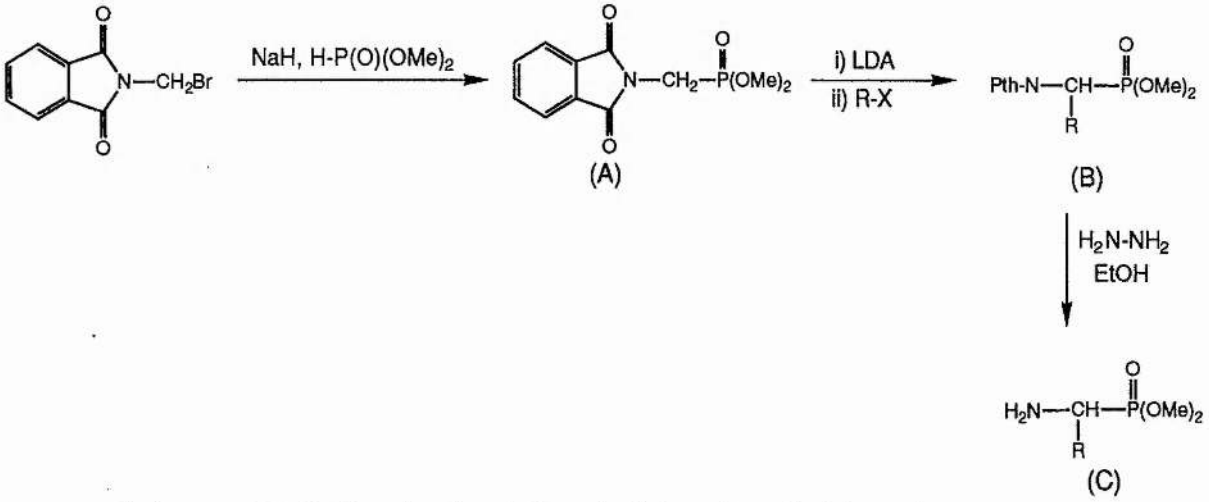
Scheme 2.3.4: Possible syntheses of p-methoxy phenylacetaldehyde

The difficulty with the third route, metallating a *p*-haloanisole to generate a nucleophile, is that the position of lithiation or formation of the Grignard migrates around the phenyl ring so as to become *ortho* to the methoxy group. The metallations are reversible, so they give the most stable form of the metallated ring. The most stable form occurs when the metal of the carbon-metal bond of the anion is able to interact with the electron-donating substituent *i.e.* is *ortho* to the substituent. Migration of metallation so that it becomes *ortho* to an electron-donating substituent is commonly observed in similar systems.[345] Generation of either the Grignard or the phenyllithium species, followed by quenching with ^2H_2O clearly showed (by proton n.m.r. spectroscopy) that deuterium was incorporated *ortho* to the methoxy group.

Therefore, the anion had undergone migration from its original *para* position so as to become *ortho* to the methoxy group. No clean reaction could be observed with the Friedel-Crafts route. Both routes were therefore set aside and new methods examined.

An initially favourable route identified was the reaction of commercially available *para*-methoxy styrene with thallium (III) nitrate[346] in methanol to give *para*-methoxy phenylacetaldehyde in one step. However, to scale up the reaction to the 50-100 mmol scale required would require the handling of large quantities of toxic thallium (III) nitrate and result in large amounts of toxic thallium (I) by-products. It was considered that metallating *p*-methoxy benzyl bromide and reacting the resulting anion with methyl formate or formyl chloride may have provided a workable route but this was not attempted. In conclusion the difficulties encountered in the synthesis of *para*-methoxy phenylacetaldehyde proved insuperable and other approaches to the phospho analogue of tyrosine were sought.

The three component condensation (Scheme 2.3.2) has two major disadvantages. It depends upon the aldehyde being readily available or easily prepared and it does not work with formaldehyde. A variation has been reported for formaldehyde[335] but this was found to be unreliable in our hands. Thus another approach to the 1-aminoalkylphosphonates was required, and a route that solves both of these problems is outlined below (Scheme 2.3.5).



Scheme 2.3.5: Synthesis of dimethyl 1-aminoalkylphosphonates

The compound A in this scheme is phthalimido protected dimethyl 1-aminomethane

phosphonate. This can then be converted to its anion with a strong base and alkylated with any required side-chain. The use of LDA or other such bases (such as LHMDS) is to be preferred for this reaction over *n*BuLi, as *n*BuLi could demethylate the phosphonate in preference to deprotonation. Phthalimido N-protection seemed especially suitable for this reaction, as there is no acidic amide proton. It was expected that this proton would be removed by the base and the resulting anion on nitrogen would suppress anion formation or alkylation on carbon. However, the work of Steglich[347] has shown that this is not so, and alkylation on carbon α to an amide NH proceeds in good yield. Treatment of intermediate (B) with ethanolic hydrazine should provide dimethyl 1-aminoalkyl phosphonates. Similar treatment of intermediate (A) would give dimethyl 1-aminomethane phosphonate.

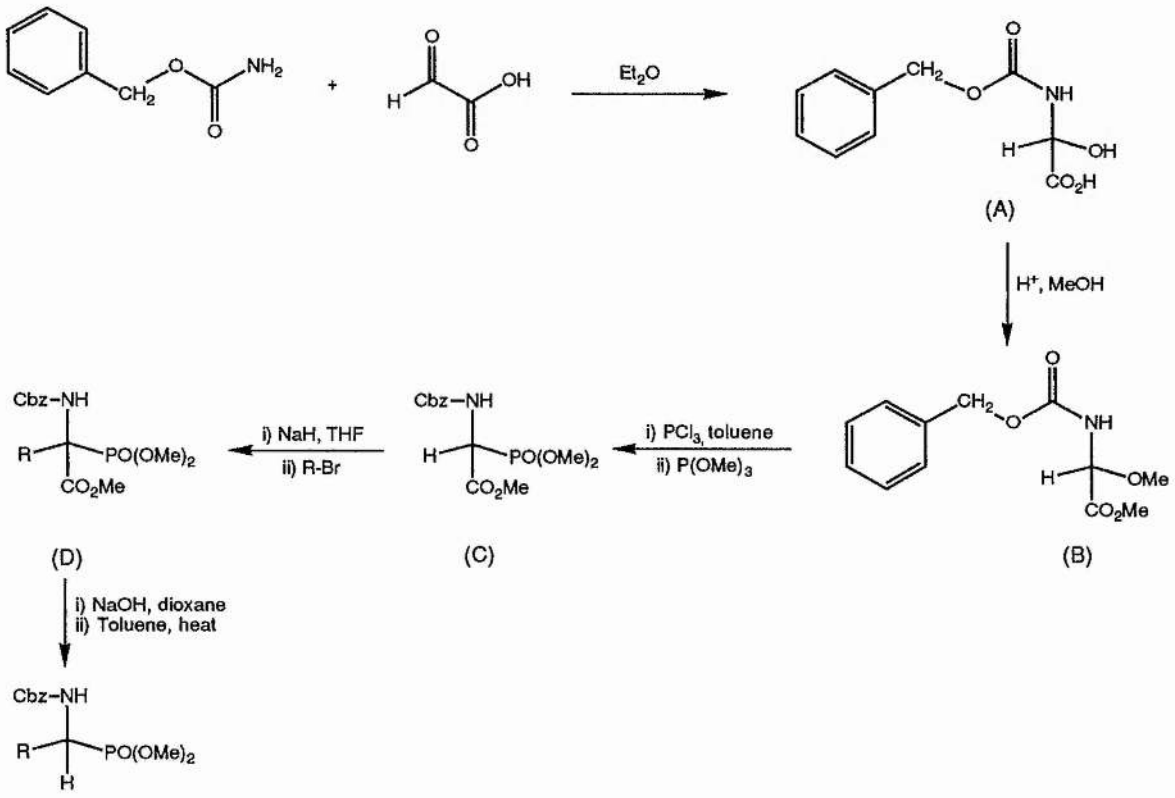
Surprisingly, the first step in Scheme 2.3.5 did not proceed if it was performed as for the ethyl analogue.[323] Treatment of N-bromomethyl phthalimide with neat trimethyl phosphite gave no detectable reaction by t.l.c., in contrast to the vigorous reaction observed with triethyl phosphite, where ethyl bromide is rapidly evolved from the reaction. Several variations on this method were attempted, using longer reaction times and elevated temperatures, without success. The literature method for the methyl analogue[348] is long and was found to give only moderate yields (45-55%). Thus, a new method was developed, as outlined above, using the anion of dimethyl phosphite, generated with sodium hydride. Isolated yields of compound A from this route approached 85%.

When this route was applied to the synthesis of dimethyl 1-aminobenzylphosphonate very low yields of product were formed. Quenching the alkylation reaction with deuterium oxide, after addition of LDA, to assess the amount of anion formation gave low levels of deuterium incorporation, as judged by proton n.m.r. spectroscopy. However, on addition of LDA to the intermediate (B) there was almost instant formation of a dark orange colour in the reaction, consistent with extensive anion formation. The low level of deuterium incorporation, however, indicates that low levels of metallation were achieved. The reason for this is unclear.

A similar approach to that in Scheme 2.3.5 was tried, using a phosphomalonate analogue, compound (C) below (Scheme 2.3.6). The presence of a carboxyl ester and a phosphonate diester bonded to the same methylene group should result in the methylene protons being very acidic, by analogy with malonate diesters. Thus, formation of an anion at this centre should be favoured and it was considered that

subsequent alkylation would be straightforward. Once the alkylated product has been formed, decarboxylative saponification (a facile process in the malonates) would give the N-protected dimethyl 1-aminoalkyl phosphonate.

All the steps, except for the decarboxylation, are from the literature, the first two being straightforward and reasonably high-yielding (around 50% over the two steps).[349] It has also been shown that the construction and alkylation of the phosphomalonate species is reasonably simple.[350] A three-component coupling reaction was attempted to form the phosphomalonate (C) directly in one step, from glyoxylic acid, benzyl carbamate and either trimethyl or triphenyl phosphite. This was unsuccessful in either case, as no product could be isolated from the reaction.



Scheme 2.3.6: Proposed route to 1-substituted dimethyl phosphonates via phosphomalonates

The major difficulty forseen with this route was the lack of selectivity in the base hydrolysis between the carboxylic and phosphonate esters. It was expected that the phosphonate ester will be hydrolysed first, due to its lower first pK_a. If the phosphonate ester is hydrolysed (and is therefore anionic) before the carboxylic

ester then there will be no stabilisation of the negative charge, α to the phosphonate, resulting from decarboxylation. Thus, decarboxylation will not be favoured. It has been found that it was possible to hydrolyse the carboxylic ester selectively in the presence of the dimethyl phosphonate.[350a] Decarboxylation by heating with 1 equivalent of triethylamine in toluene should then proceed smoothly.

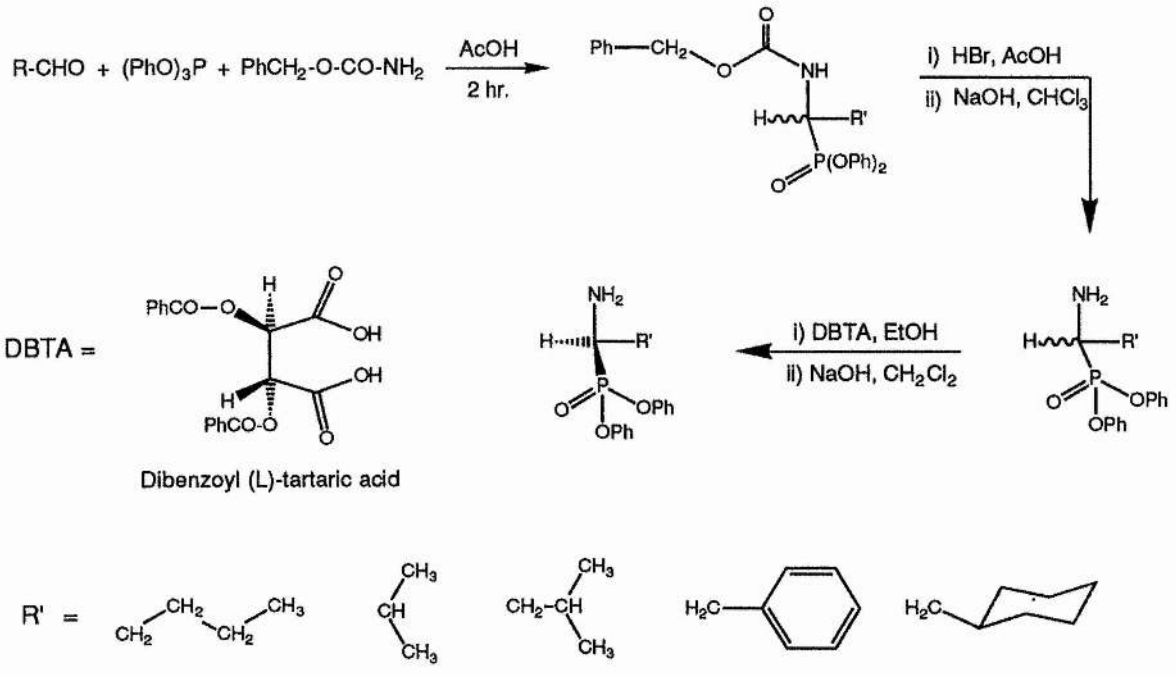
However, the alkylation step gave several unidentified by-products and no product. Quenching the anion (generated with sodium hydride) with ^2H_2O showed a low (up to 30-40% of theoretical) incorporation of deuterium at the methylene group, as judged by proton n.m.r. spectroscopy. This indicated a low degree of anion formation, which could not be significantly increased by various alterations to the reaction time and temperature. The insuperability of this problem led to the route being abandoned. Attention was then turned to stereoselective routes.

2.3.2 Stereoselective routes to aminophosphonic acids

Clearly a stereoselective route to optically active α-aminophosphonic acids would be very useful and many methods are available (reviewed in[351]). Addition of phosphites to chiral imines was the first method used.[352] Some approaches have alkylated chiral phosphonamidates[353] or chiral aminomethylphosphonates,[353] whereas other approaches have relied upon alkylating aminomethylphosphonates with attached chiral auxiliaries (a variant on the glycine enolate approach).[355] Addition of phosphites to nitrones[356] or oxoiminium salts[357] and the cycloaddition of ethylene to a chiral nitrone[358] and subsequent reaction with a phosphite have also been utilised. A "reverse" protocol involving the addition of chiral enamines to acyliminophosphonates[359] has been useful only for the production of non-proteinogenic amino acid analogues. A few enantiomerically pure 1-aminoalkyl phosphonates have been synthesised by a chiral gold (1) complex catalysed asymmetric aldol condensation with isocyanomethyl phosphonates.[360] Electrophilic amination of chiral phosphorus-stabilised anions has given good yields and high selectivity.[361] There is one report on the asymmetric hydrogenation of dehydrophosphonic acid derivatives.[362] Lewis acid promoted cleavage of homochiral acetals by alkyl phosphites followed by stereoselective amination has

also been used.[363]

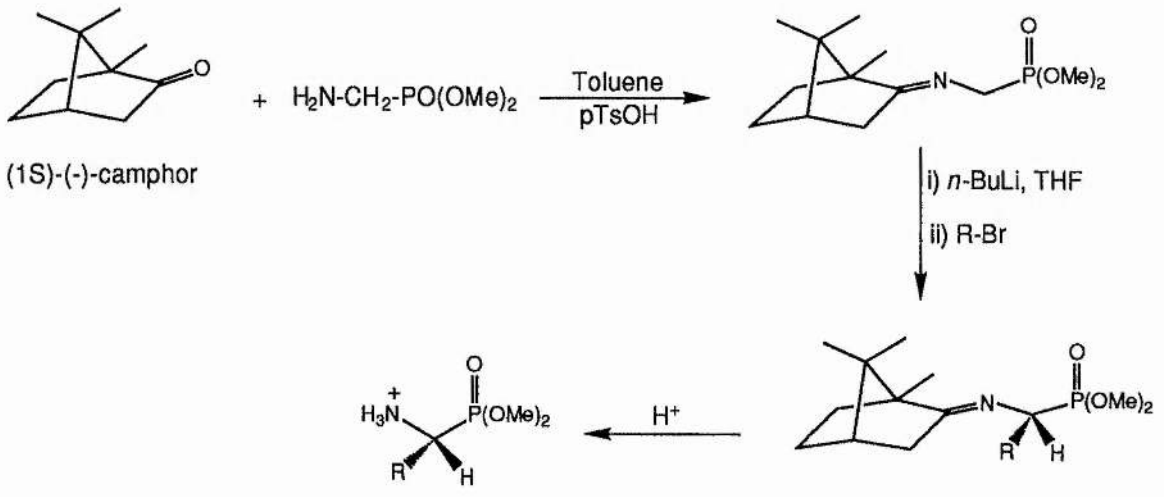
Another possible route to stereodefined phosphonamidates is to resolve racemates produced, for example, by the three component coupling (Scheme 2.3.1). It is possible either to resolve using a chiral base, to react with the deprotected phosphonic acid, or a chiral acid to react with the deblocked amino group. It was decided, following literature precedent,[364] that the simplest route would be to use a chiral acid. In this approach (Scheme 2.3.7) the racemic 1-aminoalkyl diphenyl phosphonate is made to react with dibenzoyl-(L)-tartaric acid to give two diastereomeric salts. These can be resolved by fractional crystallisation and the enantiomerically pure 1-aminoalkyl phosphonate released by treatment with sodium hydroxide and extraction with dichloromethane.[365] Resolution can also be achieved with dibenzoyl tartaric anhydride, by fractionally crystallising the diastereomeric imides produced from the reaction with the 1-aminoalkylphosphonate and releasing the resolved 1-aminoalkylphosphonate by acid hydrolysis.[366] A similar approach uses penicillin acylase, which only hydrolyses the (R) enantiomer ((L)-amino acid analogue) of *N*-acylated 1-aminoalkylphosphonates.[367]



Scheme 2.3.7: The synthesis and resolution of the 1-aminoalkyl diphenyl phosphonates

The two step removal of the Cbz group was necessitated by the great difficulty found with hydrogenation. The reaction proceeded very slowly, even under pressure and with high proportions of catalyst (up to 30%) in a range of different solvents (methanol, ethanol and ethyl acetate). The product also proved impossible to isolate in good yield. Eventually it was found that acetic acid was required in the hydrogenation to give a reliable reaction, but even then isolation of the product in a pure form was difficult. In contrast the two step method gave good yields (75% over two steps) of pure product and could be carried out quickly. Resolution was accomplished using the diphenyl phosphonates as the dimethyl esters are not stable to the HBr/AcOH treatment. Thus, two steps are required after the resolution, to give the half ester suitable for coupling. Resolution was only successfully carried out for the benzyl analogue.

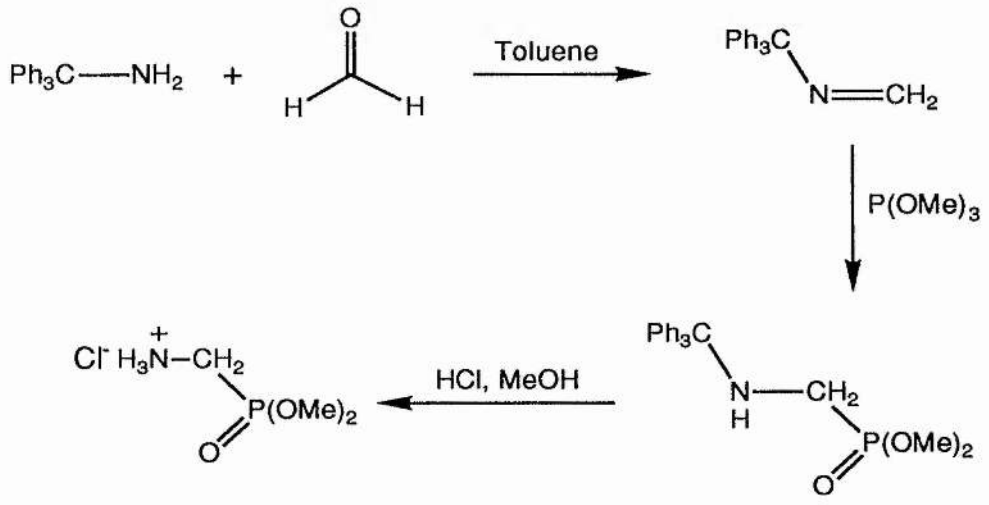
The alkylation of aminomethylphosphonates with attached chiral auxiliaries was examined as a complement to this method, *i.e.* for those compounds for which the requisite aldehyde for the three component coupling is not readily available (as above). A glycine enolate approach was chosen in which a chiral auxiliary, camphor, and 1-aminomethanephosphonate are condensed to form a chiral imine, which can then be easily alkylated (Scheme 2.3.8). Camphor was selected as the chiral auxiliary due to its easy availability and the high enantiomeric excesses reported by Schollkopf[368] for the alkylation.



Scheme 2.3.8: Schollkopf's chiral synthesis of 1-aminoalkylphosphonates

This method is essentially a stereoselective version of the routes shown in Schemes 2.3.5 and 2.3.6

The chiral synthesis required an efficient route for synthesising the dimethyl aminomethylphosphonate. The variation on the three component coupling with formaldehyde[335] was found to be unreliable in our hands and so other methods were examined. The route based on Scheme 2.3.5, deprotecting the phthalimido protected dimethyl aminomethylphosphonate with ethanolic hydrazine, was also examined and rejected as the product proved difficult to purify. A synthesis based on the addition of phosphites to triphenylmethanimine was found to be reliable and high yielding (Scheme 2.3.9).[369] The readily available triphenylmethylamine and aqueous formaldehyde are condensed together to give the imine in around 80% yield. Reaction with trimethyl phosphite gave the trityl protected dimethyl aminomethylphosphonate. Treatment with dry HCl gas in methanol gave dimethyl aminomethylphosphonate hydrochloride, suitable for condensation with camphor.

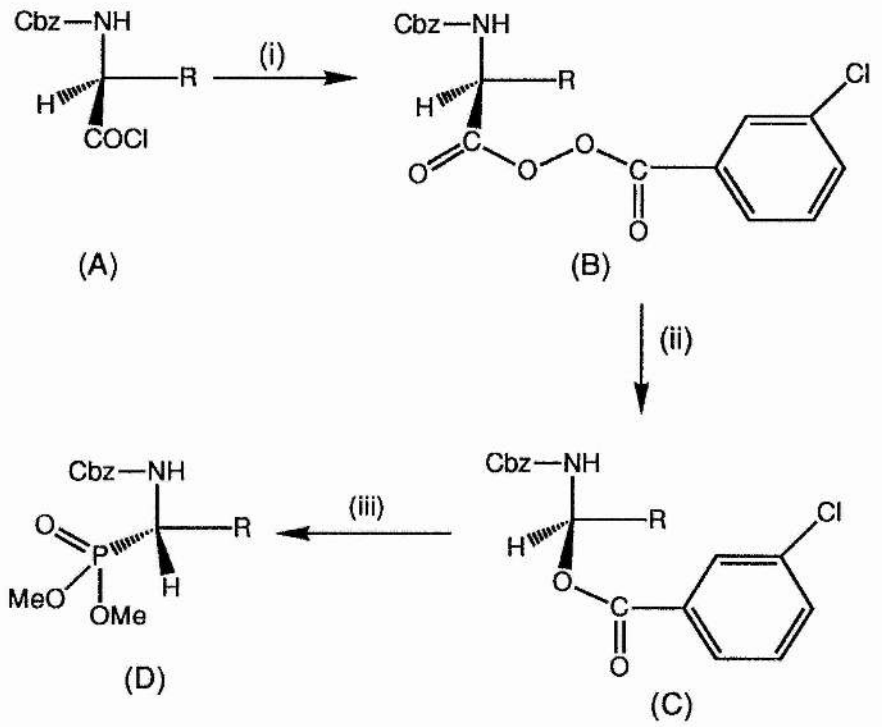


Scheme 2.3.9: Synthesis of dimethyl aminomethylphosphonate hydrochloride

However, when the condensation reaction was tried very small amounts of the imine were formed (less than 10% overall yield).

Another stereoselective route to phosphoamino acids was investigated to overcome this problem. In this route the starting point is a N-protected α-amino acid, which is transformed into the *m*CPBA peranhydride (B), *via* the acid chloride (A). This

would then be thermally decarboxylated with retention of configuration at the α centre,[370] to give the *m*-chloro benzoyl ester (C). This could then be made to react with dimethyl phosphite (or its anion) with inversion at the α centre to give the phosphoamino acid diester. To give phospho analogues of the L-amino acids the D-peranhydride would be required. This is easily available from the D-amino acid. Preliminary investigations on this route indicated that yields were low for the formation of the peranhydride.



Scheme 2.3.10: Alternative stereoselective synthesis of phosphoamino acids

Reagents: (i) *m*CPBA, Et_2O, pyridine. (ii) Toluene, reflux. (iii) $HPO(OMe)_2$, toluene.

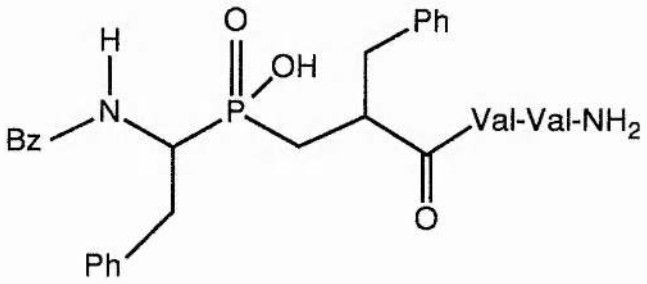
2.4 Phosphorus based protease inhibitors

Phosphorus based inhibitors have also been developed for other classes of proteases and other aspartic proteases. The work of Powers[371] has shown that serine proteases are selectively and covalently inhibited by (α-aminoalkyl) diphenyl

phosphonates. Phosphonates and phosphonamidates show non-covalent but still potent inhibition of the metalloproteases thermolysin[372] and carboxypeptidase A.[373] Phosphonates were also found to be potent inhibitors of pepsin and penicillopepsin.[374] Bartlett[375] has produced phosphinate based inhibitors for pepsin that were among the most potent yet found for the enzyme.

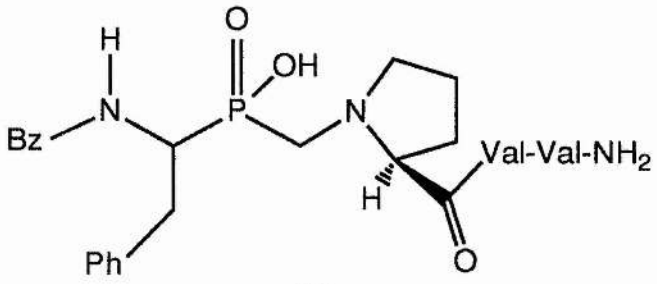
2.4.1 Phosphorus based inhibitors for the HIV-1 protease

The use of phosphorus based inhibitors for the HIV-1 protease has been investigated by a number of groups. Grobelny *et al.*[376] demonstrated that tetrahedral phosphinates incorporated into peptides capable of minimally spanning P_2-P_3' are effective *in vitro* as highly potent inhibitors of HIV-1 protease. Similar conclusions were arrived at by Dreyer *et al.*[377] Affinity for the enzyme was found to be quite sensitive to substitution at the P_1 and P_1' sites, with the Phe-Phe analogue providing superior binding to its Phe-Pro counterpart (see compounds (A) and (B)). Analogues containing a Phe-Pro mimic were poor inhibitors (see below). The binding affinity of phosphinate inhibitors was also sensitive to pH; a 500-fold increase in affinity was observed for some compounds as the pH was lowered from 6.5 to 4.5 (see below).



(A)

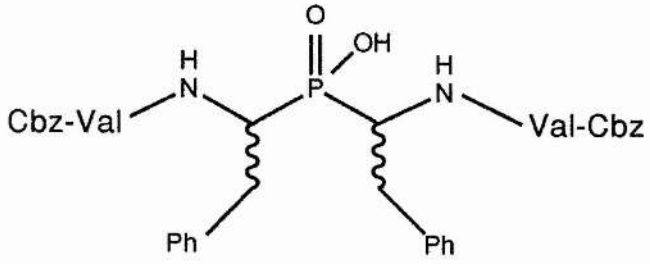
Compound (A) possesses a K_i value of 0.6 nM at pH 4.5.



(B)

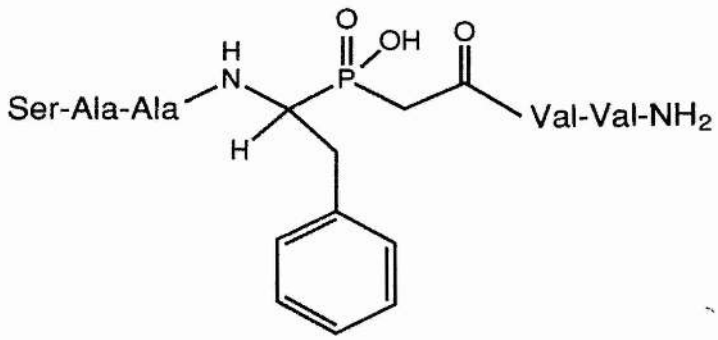
Compound (B) possesses a K_i value of 450 nM at pH 4.5. A similar compound (the so-called "exploding transition state analogue") has recently been synthesised by Ikeda *et al.*[378] and has a K_i of 8.2 nM at pH 6.2.

The HIV-1 protease is a C_2-symmetric dimer, (see Section 1.6.2.2) and this symmetry has been exploited in the design of symmetrical phosphorus-based inhibitors (see Section 1.9.2). Recently, Peyman *et al.*[379] synthesised the first symmetrical phosphinic acid-based inhibitors for the HIV-1 protease. Compound (C), as a mixture of stereoisomers, gave an IC_{50} value of 36 nM at pH 5.5.



(C)

Whilst optimising symmetry, compound (C) in fact has the disadvantage of positioning the benzyl group at P_1' incorrectly. Compound (A) is a more accurate substrate analogue, despite the fact that a carbon atom replaces the nitrogen present at P_1' in the substrate. Compound (B) also positions Pro at P_1' incorrectly. This incorrect positioning of some of the binding determinants within the active site may cause a decrease in binding affinity. Hence, compound (C) is more active than compound (A) which is in turn more active than compound (B). However, incorporation of a phosphinate into a different HIV inhibitor, compound (D), led to a disappointing K_i of 4.5 μM.[380]



(D)

The decreased potency at the assay pH (6.0) is thought to be due to the negative charge on the phosphinic acid. Inhibitors based on phosphinic acid have proven extremely potent against pepsin, where the assay pH is 3.5 and the neutral, protonated form of the phosphinic acid is predominant. It is proposed that there is an unfavourable electrostatic interaction between the negatively charged phosphinate and the aspartate carboxylate at the active site, which reduces inhibitory potency. This theory is supported by the decrease in the K_i of compound (D) to 120 nM at pH 3.6. Also, penicillopepsin is much less inhibited by this class of inhibitor than pepsin, due to the higher pH optimum of penicillopepsin, which means that the preponderant form of the inhibitor at the assay pH is anionic. The inhibition of penicillopepsin also shows pH dependency, increasing with decreasing pH, supporting the hypothesis of the neutral phosphinic acid as the active species.[375b,381] A similar phenomenon is observed with phosphinate inhibitors of renin.[382] Compound (B) and its analogue (see above), however, both have very low K_i's at high pH, possibly indicating that their inhibition is not very pH dependent.

Literature precedent therefore seemed to indicate that the phosphinate/phosphonamidate based inhibitors would be poor inhibitors, due to the adverse charge interaction with the active site carboxylate. The neutral methyl esters therefore seemed to be more promising targets. Surprisingly, however, a variant on compound (C) above was found to be around 40-fold more potent as the deprotected form (presumably anionic at the assay pH) than as the ethyl ester.[383] This is supported by the observation of "surprising pH dependence" of the inhibition of a stereodefined version of compound (B).[384] The reason for this reversal is unknown. To date there has been only one report of phosphonamidate-based inhibitors for the HIV-1 protease, by McLeod *et al.*[385]

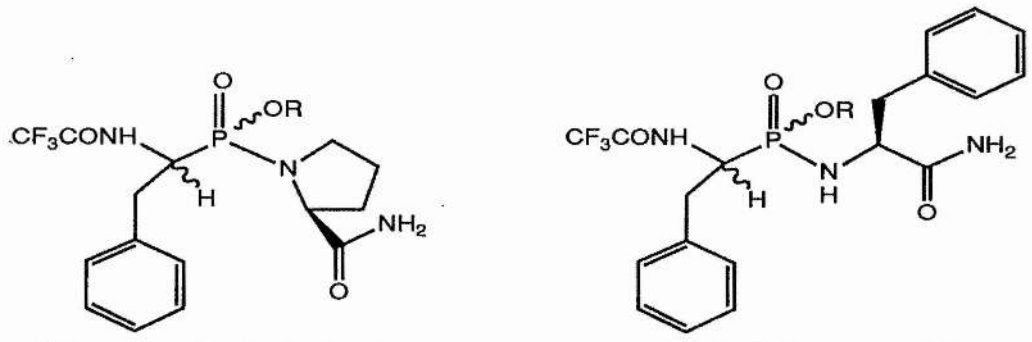


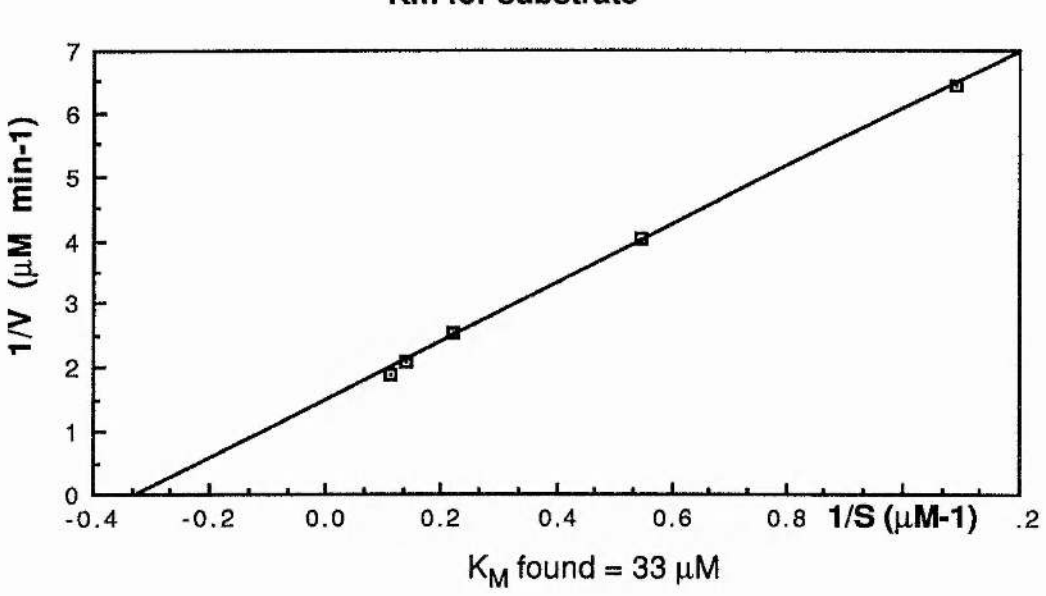
Figure 2.4.1: The phosphonamidate inhibitors of McLeod et al.[385]

where R = H or CH_3

The compounds are racemic dipeptide mimics, with IC_{50}'s *in vitro* of around 50 μM. It is interesting that these compounds are inhibitors at all, as the protease is known to require hexapeptide substrates (see Section 1.6.2.4) and the shortest inhibitors of other classes so far published are tetrapeptide mimics. Also, the difference in inhibitory potency between the neutral methyl esters and the anionic phosphonamidates is very small (the phosphonamidates have approximately a 50% higher IC_{50} than the methyl esters). With the phosphinates and phosphonates discussed above the difference is much more marked. There has also been one report of moderately potent, symmetric N-PO-N based inhibitors, with Pro at S_1 and S_1'.[386]

2.5 Results of inhibitor testing

Before beginning assays on our inhibitors it was necessary to validate our assay conditions and obtain a K_M value for the substrate used (KARV-Nle-Phe(NO_2)-EA-Nle-G-NH_2). This was performed essentially as described by Richards *et al.*[387] The assay relies on monitoring a decrease in absorbance at 300 nm, due to cleavage of the Nle-Phe(NO_2) bond. The initial velocities (at less than 20% reaction) and substrate concentrations were plotted on a Lineweaver-Burke plot, shown below.



The K_M for the substrate was determined from the slope and the intercept of the line and was found to be 33 μM under the assay conditions, which compares well with the literature value of 35 μM.[317] Thus our assay conditions were validated.

The initial inhibitors were tested using a variety of co-solvents. At first DMSO was used (at 2% final concentration) but was found to interfere with the assay as it absorbs quite strongly at 300 nm. It has also been found that DMSO acts as an inhibitor for HIV-1 protease and causes its activity to decline.[388] Some of the inhibitors were soluble enough to be assayed using 100 μM stocks in the assay buffer, but others were not. Eventually methanol (at 2% final concentration) was selected as the cosolvent of choice as it has no interference with the assay and has only a moderate effect on the enzyme activity. Once we obtain an IC_{50} value for an inhibitor we can obtain the K_i simply from the relationship:

$$K_i = IC_{50} / 1 + (S/K_M)$$

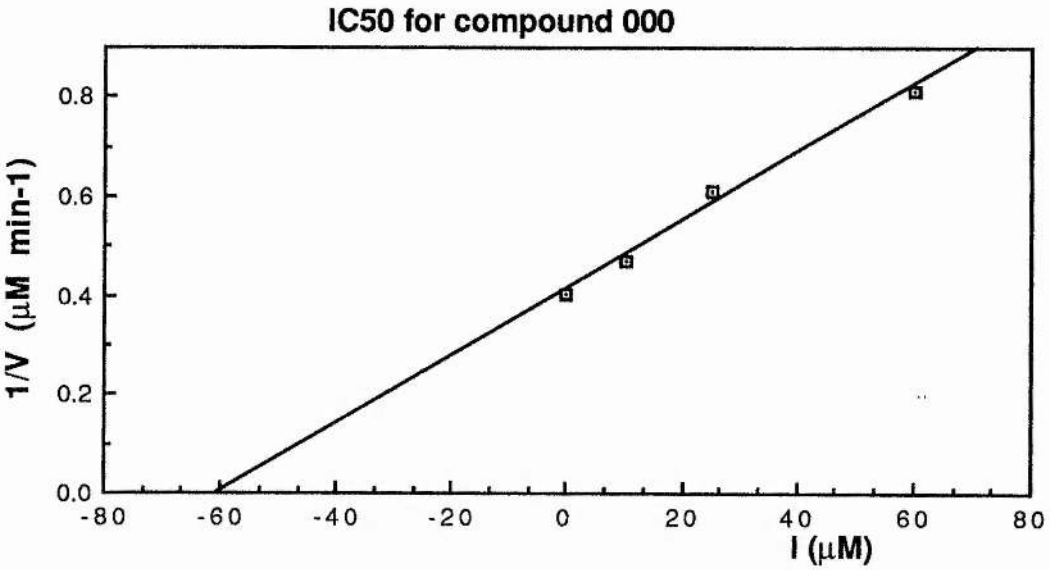
Our initial target compound 000 was prepared according to the strategy outlined above and was tested as both the methyl ester and the phosphonamidate.

Boc-N-F-ψ[PO_2H-N]-P-I-NHiBu

Compound 000

It was expected that at the assay pH (5.5) the phosphonamidate would be entirely anionic.[317] This was supported when compound 000 was tested as a mixture of two epimers at the α-carbon of the phosphophenylalanine residue and found to have an IC_{50} value of 60 μM (see graph below). This rather high value confirmed the expectation that the anionic compounds would be poor inhibitors and attention was turned to the methyl esters.

The graphical treatment of the data is based on that of Dixon.[389]



All four diastereomers of the methyl ester analogue of compound 000, compound 100, were synthesised as for compound 000, omitting the final deprotection step. They were then tested individually and the results are shown below.

Boc-N-F-ψ[PO(OCH_3)-N]-P-I-NHiBu

Compound 100

Table 2.1: *Test results for compound 100*

Code no.	Epimer at (C,P)	IC_{50} (μM) *in vitro*	*in vivo*
1001	(R, R)	20	45
1002	(R, S)	140	>100
1003	(S, R)	300	N.T.
1004	(S, S)	73	N.T.

The deprotected version of the compound was found to be a poorer inhibitor than the ester, as predicted above. This is probably due to the repulsive interaction between the anionic phosphonamidate and the active site aspartate, see above. However, the difference in the potencies is much smaller than would be expected from the literature values quoted above, where the differences are often greater than an order of magnitude. The reason for this is unclear, and molecular modelling has been unsuccessful in assisting the development of a hypothesis. These results are, however, entirely consistent with those reported (see above) by McLeod *et al.*[385] where a very small difference in the inhibitory potency of phosphonamidates and their methyl esters was also observed.

The relative inhibitory potencies of the compounds show some interesting trends. As expected the (R) configuration at the α-carbon of the phosphophenylalanine residue gives the more potent inhibitors (compare 1001 with 1003 and 1004). It is known that peptidic inhibitors of the HIV-1 protease having D-configured amino acids (equivalent to the (S) configuration of the phosphophenylalanine residue) at P_1' are poor inhibitors for the HIV-1 protease. For example, Cushman *et al.*[390] found that D-Phe at P_1' increases IC_{50} by 11 fold over the L-Phe epimer in a series of hydroxyethylene inhibitors. It is known that for the mammalian aspartic proteases if one of the residues flanking the scissile bond (P_1 or P_1') is of the (D) configuration then the scissile bond is rendered completely resistant to cleavage[391] *i.e.* the peptide is not bound.

The variable effect of the configuration at phosphorus on the IC_{50} is more difficult to understand. Comparison of 1001 with 1002 and 1003 with 1004 shows two opposite effects. In the first case the change from (R) to (S) configuration increases the IC_{50}

and in the second the same change decreases the IC_{50}. This difference has been partly rationalised using molecular modelling (see Section 2.5.1).

The IC_{50} values obtained are somewhat higher than expected for a hexapeptide inhibitor. A program of modification was undertaken to explore the effect of the P_3 residue on binding, and a panel of compounds were synthesised that lacked a P_3 group. The P_3 group was chosen for removal as molecular modelling (see Section 2.5.1) indicated that it may have little effect on inhibitory potency. Compound 200 was chosen as the target for this experiment and was synthesised and resolved by Nicholas Camp as outlined in Scheme 2.2.1.

Cbz-F-ψ[PO(OCH_3)-N]-P-I-NHiBu

Compound 200

In compound 200 the P_2 asparagine residue in compound 100 is replaced by the Cbz protecting group and the P_3 group is deleted. The Cbz group was chosen as it is a good analogue of the benzyl group that is often found at P_2 in many potent inhibitors. These compounds were tested in the *in vivo* assay only and the results are shown below.

Table 2.2: *Test results for compound 200*

Code No.	Sequence	Epimer at (C,P)	IC_{50} (μM) *in vitro*	IC_{50} (μM) *in vivo*
2001	CbzF-ψ–P-I-NHiBu	(R,R)	N/T	3
2002	"	(R,S)	N/T	25
2003	"	(S,R)	N/T	100
2004	"	(S,S)	N/T	25

The results show that the configuration at phosphorus has a variable influence on the IC_{50}, compare 2001 with 2002 and 2003 with 2004. In the first case a switch from (S) to (R) increases the IC_{50}, in the second the same change decreases the IC_{50}. This is consistent with the results obtained from compound 100. The basis for this effect probably lies in the cooperativity between the S_1 binding site and the active

site diad of aspartates. The equivalence of the IC_{50} values for 2002 and 2004 is also difficult to understand, as the configuration at the α-carbon of the phosphophenylalanine residue usually has a large influence on the potency of an inhibitor (see above). Strikingly, the loss of the P_3 residue appears to have had essentially no effect on the inhibitory potency of the compounds.

In common with other studies on HIV-1 protease inhibitors (see Section 1.6.3.2) we explored compounds based on pseudo-symmetric scissile bond analogues. The Phe-ψ[$PO(OCH_3)$-NH]-Phe and Cha-ψ[$PO(OCH_3)$-NH]-Phe (where Cha represents cyclohexylalanine) analogues were investigated. A variety of unresolved, epimeric inhibitors of the sequence of compound 300 (see below) were synthesised, based on these scissile bond mimics. The results from compound 200 were also taken into account, so that compounds with and without a P_3 residue were synthesised.

X-ψ[$PO(OCH_3)$-NH]-Phe-Ile-NHiBu

Compound 300

Table 2.3: *Test results for compound 300*

Code No.	Sequence of X	Epimer at (C,P)	IC_{50} (μM) *in vitro*	*in vivo*
3001	Cbz-F	mixture	2	2
3002	Cbz-Cha	mixture	1	5
3003	Boc-N-F	mixture	2	N/T
3004	Boc-N-Cha	mixture	4	N/T

These compounds are mixtures of 4 diastereomers and the diastereomers are not produced in equal amounts. Therefore, the IC_{50} of the most active diastereomer could be less than 25% of the IC_{50} for the mixture, if that diastereomer only shows significant activity. Once again the presence or absence of a P_3 group appears to have almost no influence on the inhibitory potency (compare 3001 with 3003 and 3002 with 3004).

As has been observed frequently with other systems the pseudo-symmetric

inhibitors are more potent than the non-symmetric analogues. The increased potency of these pseudo-symmetric inhibitors is mirrored by the lower K_M and higher k_{cat}/K_M for substrates containing an Aromatic * Aromatic scissile bond compared with those containing an Aromatic * Proline scissile bond.[207,392] This is shown in the kinetic parameters for the following two substrates, which differ only in their P_1' residue:

S-Q-N-**Tyr** * **Pro**-I-V	K_M 2.3 mM	k_{cat}/K_M 60.9 $mM^{-1}min^{-1}$
S-Q-N-**Tyr** * **Tyr**-I-V	K_M 0.8 mM	k_{cat}/K_M 91.2 $mM^{-1}min^{-1}$

However, the Aromatic * Proline scissile bond analogue does not always result in a decreased inhibitory potency compared with the pseudo-symmetric Aromatic * Aromatic analogue.[220] Hydroxyethylamine based inhibitors with Aromatic * Proline scissile bond mimics are often equipotent with, or more potent than, the equivalent Aromatic * Aromatic version. This may be due to the water-bridged hydrogen bond to the flaps of the protease (see Section 1.6.4), pulling the ends of the P_1 and P_1' residues down away from the active site aspartates and pushing the Pro residue up into the S_1' pocket, filling it more fully. However, this will only occur when there is sufficient distance between the hydrogen bond donors in the inhibitor, to allow the peptide backbone to adopt a more curved structure. Thus, this effect is observed with the hydroxyethylamine compounds, but not with hydroxyethylene compounds, which have one atom less between the hydrogen bond donors involved in bonding to the water molecule.

Having observed the rather small influence of the P_3 group on inhibition we attempted to determine the optimal P_3 side-chain to determine if the requirements for this side-chain were the same as for other classes of inhibitors. Literature precedent[220] indicated that a favourable side-chain at this position was the quinoline-2-carbonyl (Qua) moiety (see Figure 2.5.1).

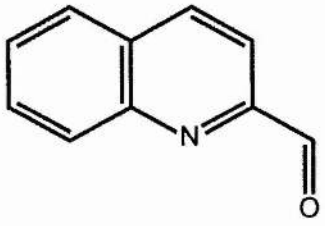


Figure 2.5.1: The quinoline-2-carbonyl (Qua) moiety

Thus a compound of the following sequence was prepared as a mixture of epimers at C-1 of the phosphophenylalanine residue and phosphorus:

Qua-N-F-ψ[PO(OCH_3)-N]-F-I-NHiBu

Compound 400

This compound was found to have an IC_{50} of 37 μM in the *in vitro* assay and 50 μM in the *in vivo* assay. A similar compound with a Cha-Pro scissile bond mimic was also prepared and a single diastereomer ((R), (R)) tested which was found to give an IC_{50} of 15μM in the *in vitro* assay. These results are consistent with those found for the compounds 100 and 300, where it was shown that the P_3 residue has little effect on the inhibitory potency. However, the quinoline-2-carbonyl moiety increased the IC_{50}, rather than decreasing it as was expected from literature precedent. Clearly the nature of the optimal P_3 residue is different for the phosphonamidates than for other classes of inhibitor. The reason for this is not understood.

The most striking observation from all these compounds is the near equivalence of the IC_{50} *in vitro* and the IC_{50} *in vivo*. This is almost entirely unprecedented and runs counter to many other results of *in vivo* testing of HIV-1 protease inhibitors and inhibitors of many other enzymes. In the case of other protease inhibitors the ratio of *in vivo* to *in vitro* IC_{50} is usually around 10-100.[393] In the case of our compounds the ratio is around 1:1 to 1:5 (see above). Therefore, either the effective concentration of the inhibitor in the cells is the same as the concentration in the medium or the compound is broken down in the cell to give a potent inhibitor of some aspect of viral metabolism. It is unlikely that the internal and external concentrations are nearly identical for two reasons. Firstly, these inhibitors probably will not easily be able to

diffuse across the lipid membrane to enter the cell. Secondly, as all exogenous compounds, particularly peptidic ones, are metabolised to some extent in the cell (in this case by host proteases) the concentration of the inhibitor inside the cell will fall steadily with time. Thus, it is possible that the effective inhibitor is some metabolic by-product of the target compound.

2.5.1 Molecular modelling of inhibitor-protease interactions

The basis for much of this work was the structure of the reduced peptide inhibitor MVT-101 complexed with the protease.[261] MVT-101 was chosen as it is highly congruent to the substrate and consequently to our inhibitor. The hydroxyethylene inhibitors are also congruent to the substrate but the hydroxy group is hydrogen bonded to the active site aspartates. This may induce a bias into the starting active site structure that may affect the outcome of subsequent calculations. The sequence of inhibitor 100 was built onto the peptide backbone of MVT-101 (see Figure 2.5.2).

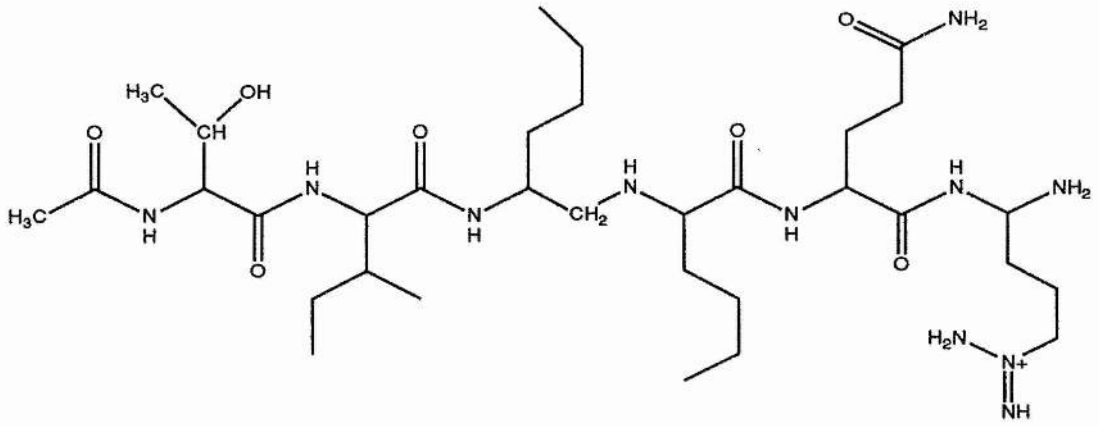


Figure 2.5.2: MVT-101

This gave an approximation of the bound conformation of the inhibitor in the active site. A substructure minimisation in the active site of the HIV-1 protease allowed the altered side-chains to relax into the enzyme subsites and adopt reasonable bound conformations. The bound conformation of compound 1001 is shown in Figure 2.5.3.

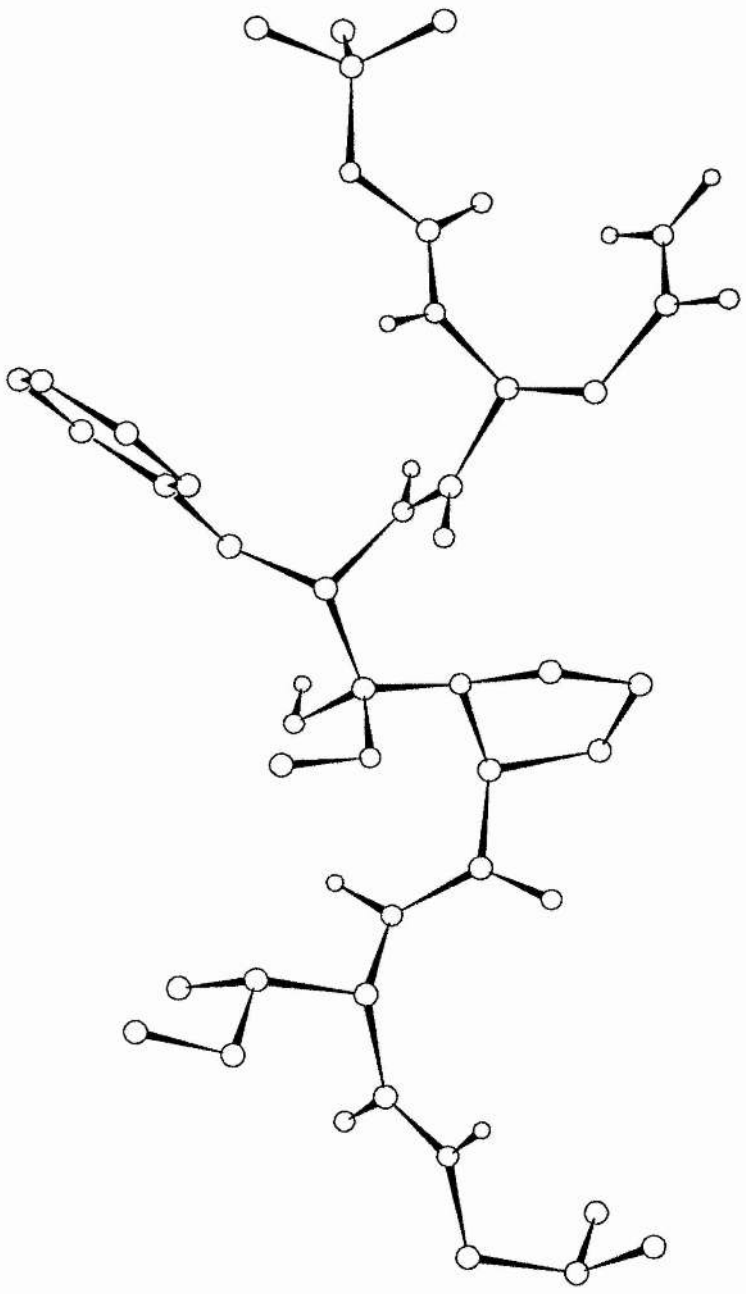


Figure 2.5.3: The modelled bound conformation of compound 1001

The interactions between the P_3 residue (the Boc group) and the residues of the S_3 pocket were examined in this model structure to identify interactions important for binding. A conformational search was then carried out on a substructure consisting of the Boc group and those residues having one or more atoms within 6 Å of the atoms of the Boc group. This was designed to identify the low energy conformations of the Boc group in the S_3 pocket and the interactions made with the selected side-chains.

A further set of calculations were carried out in which the methylene of the reduced peptide moiety was replaced by $PO(OCH_3)$, to give the full sequence of our target compound. This was carried out using the INSIGHT program to take advantage of the generalised force-field available in INSIGHT (CFF91[394,395]) that allows calculations to be carried out on systems for which high-quality bond parameters are not available. The differences in the side-chain conformations between the structures with and without phosphorus were found, as expected, to be very small (the two structures overlaid with an r.m.s. displacement of around 0.1 Å).

The number of interactions found in either case between the Boc residue and the enzyme were found to be limited. The main contacts made by the Boc group at P_3 are with residues Val 82B, Gly 48A and Arg 8B of the S_3 pocket (A represents amino acids from one of the monomers, B from the other). All these contacts are non-polar and are not expected to contribute greatly to the binding energy. The major polar interaction in the S_3 pocket, with Asp 29A, was not always observed with the Boc group, which probably reflects deficiencies in the methodology rather than an actual lack of interaction. Thus, it is not expected that this residue will add greatly to the potency of our compounds. The interactions identified compare reasonably well with the interactions identified in the published X-ray structures of inhibitors co-crystallised with the protease and are illustrated in Figure 2.5.4

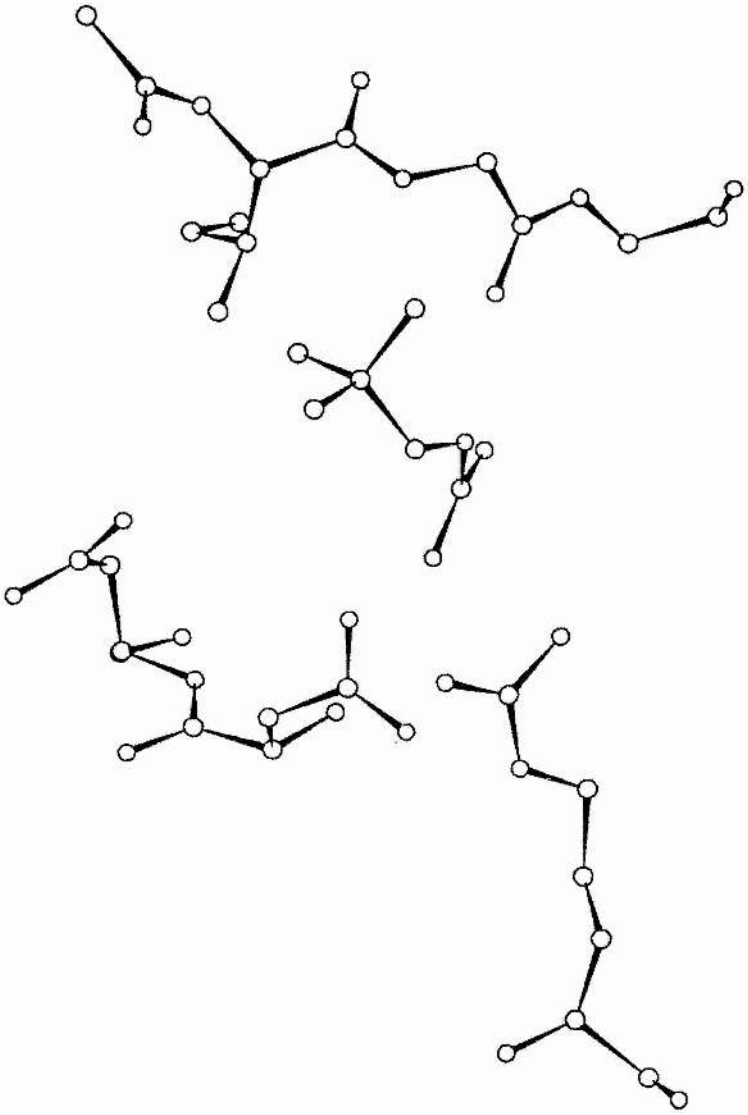


Figure 2.5.4: Interactions in the P_3 pocket

The observed improvement in the binding of the Phe-Phe and Cyc-Phe scissile bond analogue containing inhibitors over those containing Phe-Pro was investigated. The possible interactions of the different P_1' side chains with the S_1' pocket were examined using two different crystal structures. The model structure of Weber[396] which has a Tyr-Pro containing substrate was used initially. The coordinates for the structure of the protease co-crystallised with a proline-containing hydroxyethylamine inhibitor, JG-365[260] (Figure 2.5.5) were then made available.

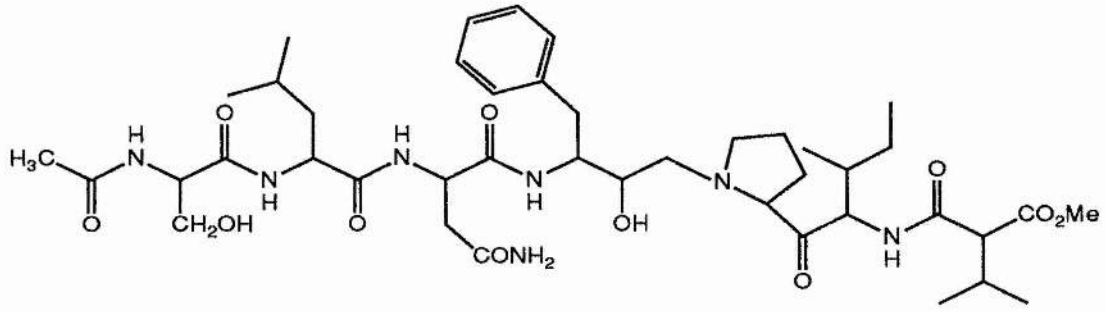


Figure 2.5.5: JG-365

The Pro residue in the model structure forms van der Waals contacts with only 3 residues, Leu 23A, Val 82A and Ile 84A and packs against Asp 25A. In the JG-365 structure the Pro residue is in contact with residues Pro 81A, Val 82A, Ile 84A and Gly 49B. In contrast the P_1 Phe side-chain in the symmetric A-74707[262] forms contacts with Leu 23A, Val 80A, Pro 81A, Ile 84A, Asp 28B and Ile 50B. Thus, proline at P_1' has less interactions with the S_1' subsite than phenyalanine at P_1', which is consistent with the usually less potent inhibition of proline-containing analogues.

This large alteration in the conformation of the active site on binding inhibitors of differing steric requirements has been observed before.[242] It should be noted that in spite of the deficiencies of the Weber model structure it is probably a better model for the interactions between our inhibitors and the enzyme than JG-365. This is due to the "bending" effect of hydroxyethylamine inhibitors discussed above, arising from water-bridged hydrogen bonding to the flap.

The procedure carried out using MVT-101 to build a model of our inhibitors was repeated for JG-365 and for the substrate Ser-Gln-Asn-Tyr-Pro-Ile-Val from the model structure of Weber. The structures were subjected to substructure minimisation and the resulting interactions compared. Essentially the same results were obtained for the interactions with the P_3 pocket. The interactions between the Pro residue and

the S_1' pocket were found to be different for the JG-365 based structure than for the model substrate structure, probably due to the bridging water mediated bending of the inhibitor backbone. Which one represents a better model for the interactions actually present in the enzyme-inhibitor complex is not known and awaits an X-ray crystal structure.

The configuration at phosphorus was investigated by building both enantiomers at phosphorus into the active site (on the inhibitor with the (R) configuration at carbon) and minimising the structures as above. Subsequently, molecular dynamics simulations were carried out on a substructure consisting of residues 22-27 of each monomer and the central residues (Phe-ψ[$POOCH_3$]-Pro) of the inhibitor in both configurations at phosphorus. It was found that the (S) configuration (see Figure 2.5.6) disrupted the coplanarity of the active site aspartates, due to the intruding methyl group. This disruption of the active site hydrogen-bonding network is an unfavourable process and is consistent with the less potent inhibition observed with the inhibitors having an (S) configured phosphorus atom.

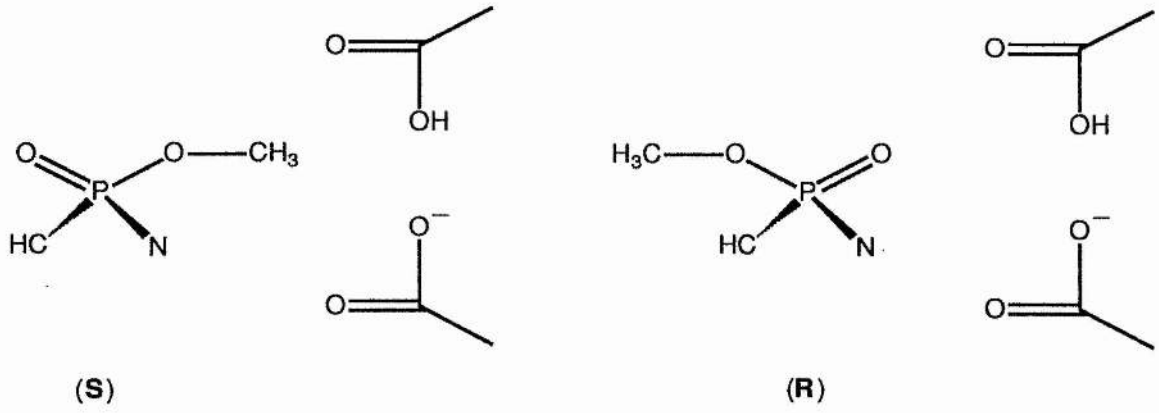


Figure 2.5.6: The epimers differing at phosphorus

The (R) configured phosphorus has the phosphoryl oxygen pointing toward the aspartates and allows them to retain their hydrogen bonding network and remain coplanar. Given that this analysis is correct it becomes difficult to account for the equivalence in binding potency of inhibitors 2002 and 2004 (see above). The explanation may lie in the extensive cooperativity of the residues of the active site, allowing 2 different configurations at phosphorus to form interactions of equivalent strength with the enzyme.

All the above work is based on the assumption that our inhibitors bind in the same way as substrates and other substrate-based inhibitors such as MVT-101. This could be demonstrated in two ways. Firstly, if it could be shown by kinetic analysis that these are competitive inhibitors then they must bind in the active site, in competition with the substrates. Alternatively, if the P_1' residue were of the D-configuration then an active-site directed inhibitor would bind more weakly, by around an order of magnitude, than an inhibitor of the L-configuration at P_1'.[390] It was also assumed that our inhibitors only have one binding mode in the active site, or, if there were two, that one represents only a small fraction of the total. This can only be demonstrated unambiguously with a highly refined X-ray crystal structure determination.

2.5.2 Time-dependency of inhibition

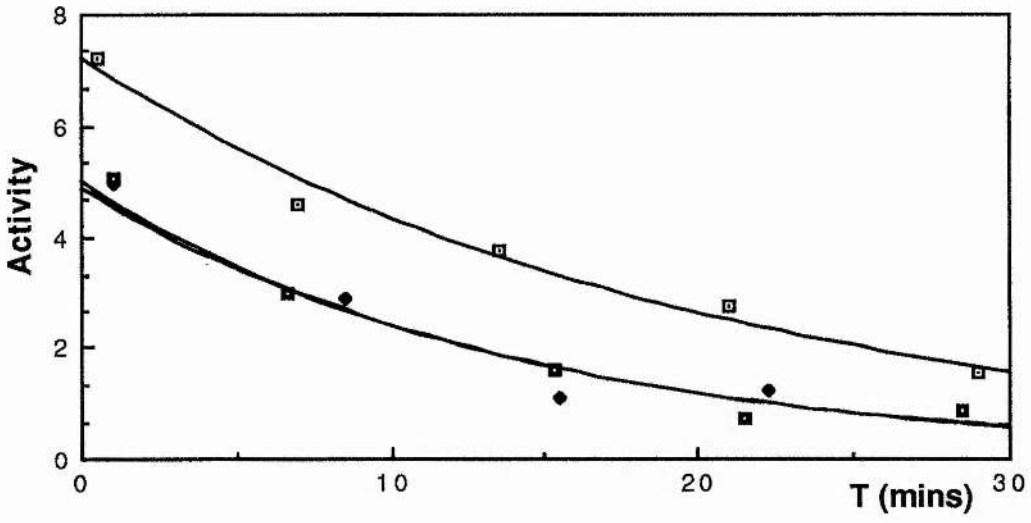
During the analysis of these inhibitors it was decided to explore the possibility of time-dependency of inhibition. Time-dependent inhibition of aspartic proteases has been observed with statine based inhibitors.[397] It has also been observed with phosphonamidate inhibitors of the metalloprotease thermolysin.[398] Rich[399] postulated that this slow binding is due to a two-stage binding process, an initial fast formation of a 'collision' complex, followed by slow isomerisation to the tightly bound form. He[399] has speculated that in the aspartic proteases this slow isomerisation is due to displacement of the tightly bound water molecule between the active site carboxyls. However, X-ray crystal structures of various aspartic proteases complexed with inhibitors show that this water molecule is lost when other, non-slow binding, inhibitors are co-crystallised with the enzyme.[129] Also, some very potent statine based inhibitors of renin show no biphasic association,[400] thus potency is also not a common feature of the slow-binding inhibitors. Equally, some potent inhibitors of renin do show biphasic binding.[401] The active site water molecule in the monomeric aspartic proteases has recently been shown not to be tightly bound at the active site.[402] Solvent isotope partitioning was used to show that the active site water molecule desorbs from the ternary (enzyme-substrate-water) complex at a rate four times higher than k_{cat}. The physical basis for this observation is unclear as inspection of a model of the substrate-enzyme interaction indicates that large

conformational changes must take place (possibly even debinding of substrate) for the water molecule to be able to leave the active site.

Fruton[67] has proposed that the slow step in aspartic protease catalysis is a conformational change, which is observed to be rapid in the cleavage of good substrates. However, as has been pointed out by Frick *et al.*[403] enzymes have minimised kinetic obstacles to conformational changes when binding substrates but not inhibitors. It is likely that the inhibitor adopts a different conformation at the active site than the substrate. Therefore, the conformational change needed to optimise interactions between the enzyme and the inhibitor will be different to, and consequently slower than, those needed to optimise the interactions between the enzyme and the inhibitor. The initial binding may be binding of the main determinants of specificity, the P_1 and P_1' sites,[171a] and the slow step may be the rest of the inhibitor side chains binding at a rate that is dependent on the conformational flexibility of the active site.

Shown below is a plot showing the loss in HIV-1 protease activity with time in the absence and presence of methanol and the presence of inhibitor 3004 (a pseudo-symmetric inhibitor) made up in a stock solution in methanol.

Time-dependency of HIV-1 protease inhibition



The black diamonds (◆) show decline in activity of the protease in the presence of 2% methanol alone. The black squares (■) show the decline in the presence of 5 μM 3004 in 2% methanol. The white squares (□) show the decline in activity of the enzyme in the absence of methanol and inhibitor. This decline may be due to

autoproteolysis.[388] Clearly there is no difference in these values, so the inhibition is not time-dependent.

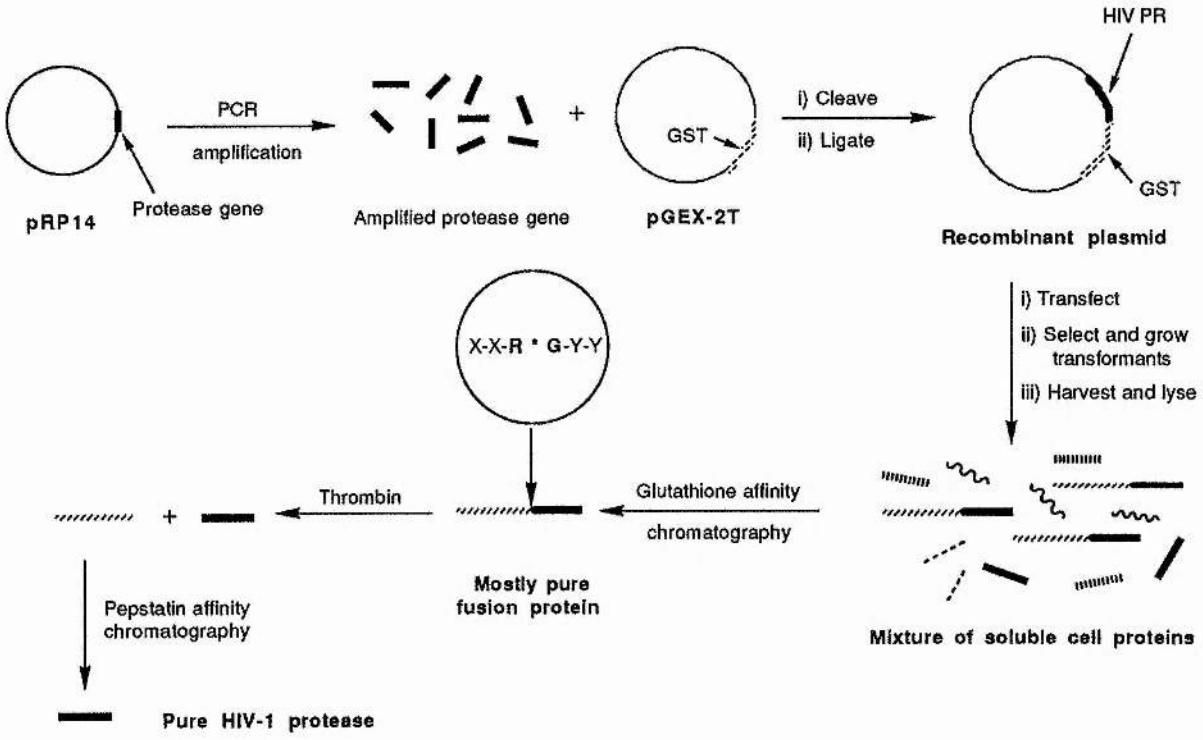
2.5.3 Further work

There are two areas that may be explored in future. Conformationally restrained analogues, with a cyclic structure at P_1 linking the side-chain to one of the phosphoryl oxygens will allow the flexibility of the S_1 pocket to be examined. Hopefully, the restraint will reduce the loss in entropy on binding and increase inhibitory potency. Compounds with glycine at P_1' will also be synthesised to examine the influence of the P_1' side-chain on inhibition and the effect of increased conformational freedom at this point.

2.6 Cloning, overexpression and purification of the HIV-1 protease

As an adjunct to the work on the inhibitors it was decided to attempt to clone and overexpress the protease in the bacterium, *E. coli*. The protease has been expressed in a number of systems, including *E. coli*,[404] insect cells[405] and yeast cells.[406] Various strategies have been used to obtain active protease, including expression of the protease as a fusion protein with non-HIV proteins[404eikl] or as part of the *gag* or *pol* precursor proteins.[404dj,406] The protease has also been chemically synthesised by a number of groups.[407]

Our strategy for obtaining the protease is outlined in Scheme 2.6.1. Firstly, the protease gene, contained in a plasmid, pRP14, was amplified, using the polymerase chain reaction (PCR) (see Section 2.6.1). The two oligonucleotide primers for this amplification contained the 6 base-pair cleavage sites for one of two restriction endonucleases (Eco R1 and Bam H1). The PCR amplified DNA is then cleaved with Bam H1 and Eco R1 and ligated into the plasmid pGEX-2T.[408] Ligation places the protease gene in frame with another protein (GST, see Section 2.6.2), under the control of an inducible promoter. The ligated plasmid is then transfected into the *E. coli* strain JM 101[409] and the protease purified from the cells after induction.



Scheme 2.6.1: Outline of strategy for the overexpression and purification of the HIV-1 protease

This PCR-based approach to obtaining recombinant proteins has been christened Expression Cassette PCR (ECPCR) by Schreiber.[410]

2.6.1 The PCR reaction

The PCR reaction (polymerase chain reaction) is a highly efficient method for selectively amplifying a DNA sequence that lies between 2 known sequences on opposite strands of double strand DNA.[411] Two oligonucleotides which are complementary to the sequences that flank the DNA of interest are used as primers for a series of synthetic reactions catalysed by a heat-stable DNA polymerase (the so-called Taq DNA polymerase). The complementary sequences are 21 b.p. long, which will ensure strong and specific binding at only the required sequence in the template DNA. The primers also have a restriction site in a non-complementary sequence that lies at one end of the complementary sequence to allow ligation of the amplified product into a suitable vector. The primers made had the sequences

shown in Figure 2.6.1. The complementary sequences are shown in bold type, the non-complementary sequences in plain type and the restriction sites are underlined. The 4 b.p. "tails" are to prevent exonucleolytic degradation during the course of the PCR process.

A. 3' AGAGGAATTC**AAAATTTAAAGTGCAACCAAT** 5'

^

Eco R1 restriction site

B. 5' ACATGGATCC**CCTCAGATCACTCTTTGCCAA** 3'

^

Bam H1 restriction site

Figure 2.6.1: The PCR primers

The PCR reaction amplifies the sequence lying between the two primers in an exponential fashion, resulting in a vast excess of the required sequence over other sequences. The basis of the amplification lies in a cycle of denaturation, annealing and extension. In the first cycle the DNA and primers mixture is heated to 94°C to denature the double helix, then cooled to 55°C to allow the primers (present in a large excess over the template DNA) to anneal to the template. Heating to 72°C allows the thermostable DNA polymerase to extend the primer as shown below. The second cycle proceeds as the first, except that there are now extended strands from the first cycle for the primers to anneal to, as well as the original template. Repetition for 25-30 cycles results in a large amount of amplified DNA.

1. **Primer 1**

««TAACCAACGTGAAATTTAAAA...
<------CCTCAGATCCTTTGCCAA----------ATTGGTTGCACTTTAAATTTT--------->
<------GGAGTCTAGGAAACGGTT----------TAACCAACGTGAAATTTAAAA-------->
...CCTCAGATCCTTTGCCAA»»

Primer 2

2.

««TAACCAACGTGAAATTTAAAA...
CCTCAGATCCTTTGCCAA----------ATTGGTTGCACTTTAAATTTT-------->
<--------GGAGTCTAGGAAACGGTT----------TAACCAACGTGAAATTTAAAA
...CCTCAGATCCTTTGCCAA»»

3.

««TAACCAACGTGAAATTTAAAA...
CCTCAGATCCTTTGCCAA---------- ATTGGTTGCACTTTAAATTTT
GGAGTCTAGGAAACGGTT----------TAACCAACGTGAAATTTAAAA
...CCTCAGATCCTTTGCCAA»»

»» - Direction of replication by the DNA polymerase.

Scheme 2.6.2: The PCR reaction

There are also competing reactions where the primers re-anneal to the longer templates, but the amount of DNA produced by these reactions only increases linearly with the number of denaturation-annealing-extension cycles. The amount of DNA from the region flanked by the primers, however, increases exponentially with the number of cycles, resulting in a DNA preparation that is essentially homogenous. This DNA was then digested with the appropriate restriction enzymes, prior to ligation into the cloning vector, pGEX-2T. The recombinant plasmid so produced was

designated pGEX-HPR. PCR from the plasmid template proceeded smoothly, but PCR from baculovirus derived *pol* DNA was very unreliable and was therefore abandoned.

2.6.2 The cloning and transfection process

The pGEX-2T plasmid (see Appendix B) contains the restriction sites for Bam H1 and Eco R1 in a short multiple cloning site (MCS). The MCS lies at the 3' end of the coding region for the glutathione-S-transferase enzyme (GST) from *S. japonicum.*[412] Cleaving the plasmid and the amplified protease gene with the same restriction endonucleases (Bam H1 and Eco R1) gives pairs of complementary ends on the plasmid and the gene. The placement of the restriction sites is such that a gene ligated into the sites will be in the same reading frame as the GST gene. Therefore, cloning the protease gene into pGEX-2T creates a fusion protein between the GST and the HIV protease, of mass 37 kDa (GST is 26 kDa, the protease 11 kDa).

There are a number of advantages to using these two restriction enzymes over any others contained in the MCS of pGEX-2T. As they have different sequence specificities and both produce overhanging ends (see Figure 2.6.2) cloning into the MCS with these enzymes will force the cloned gene to be only in one orientation and in the same reading frame as the GST.

Bam H1 - G // GATCC
CCTAG//G

Eco R1 - G//AAGGC
CTTCC//G

// - site of cleavage

Figure 2.6.2: Cleavage sites for the restriction endonucleases Bam H1 and Eco R1

Using two restriction enzymes also eliminates the need to prevent the plasmid religating without having incorporated the gene to be cloned. As the overhanging ends produced by the two enzymes are incompatible they will not base pair (the necessary step prior to ligation) in the absence of a linker with the correct complementary ends (the gene to be cloned). Thus, ligation will not occur in the absence of the gene to be cloned. Also, the enzymes have similar buffer

requirements for activity so the digestions can be performed simultaneously, rather than sequentially.

In pGEX-2T the GST gene is under the control of the *tac* promoter. The *tac* promoter is a hybrid of sequences from the *trp* and *lac* promoters.[413] It has an upstream sequence (the so-called -35 sequence) from the *trp* promoter and the sequence closer to the transcription start point (the TATA or Pribnow box) is from the *lac* UV5 promoter. As it is such a strong promoter it is necessary to keep it repressed or the gene(s) under its control will be expressed before induction *i.e.* constitutively. In pGEX-2T repression is achieved by the *lac* repressor protein, which is expressed constitutively in amplified amounts by the *lacI*^{q} gene. This represses transcription from the *tac* promoter in the absence of inducer, by binding to the promoter region. Induction is achieved with isopropyl-β-D-thiogalactoside (IPTG), a non-metabolizable analogue of galactose. Once induced with IPTG the *tac* promoter will give high levels of transcription of proteins under its control and thus high levels of expressed protein.

As the protease is fused to GST in pGEX-HPR the fusion protein is also under *tac* control. Between the 3' end of the GST gene and the MCS there are two codons coding for the cleavage site of the serine protease thrombin, Arg-Gly (see Scheme 2.2.12). Therefore, the thrombin cleavage site will be near the junction of GST and the protease in the fusion protein. Induction of expression from this promoter will result in the synthesis of a fusion protein, consisting of GST fused to the protease through the thrombin cleavage site. Thus, at a suitable stage in the purification, the protease can be released from the fusion by treatment with thrombin.

Many proteins have been isolated using GST fusion proteins.[414] The ease of binding of specific proteins to a glutathione agarose column means that the binding and elution conditions are extremely mild. A further advantage of the fusion protein protocol is that the yields of protease are often much higher than those obtained for expressing the "bare" protease.[415]

Just after the Eco R1 restriction site in pGEX-2T lie stop codons (TGA) in all three reading frames (Figure 2.6.3). Thus, transcription of the protease gene will stop, regardless of what reading frame is being transcribed, just after the end of the inserted gene. After the manipulations (ligation, expression and cleavage with thrombin) are complete the fusion protein will have the junction sequences shown in the Figure. Bases in bold type arose from the amplified protease gene, bases in

ordinary type from the plasmid pGEX-2T. Amino acids in bold type are part of the native protease sequence, those in ordinary type are from the plasmid. The numbers refer to the number of the amino acids in the protease sequence.

5' end of the protease sequence

5'...CTG GTT CCG CGT G**GA TTC CCT CAG**..... 3'
Leu Val Pro Arg Gly Ser **Pro Phe**.....
^ 1 2
Thrombin cleavage

3' end of the protease sequence

5' ...**AAT TTT G**AA TTC ATG GTG ACT GAC TGA 3'
Asn Phe Glu Phe Met Val Thr Asp STOP
98 99

Figure 2.6.3: The 5' and 3' junctions of the protease fusion
The stop codons are underlined.

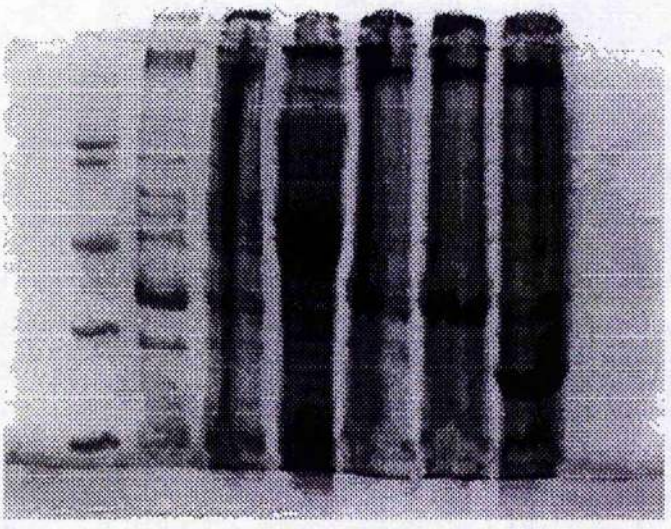
The recombinant protease will therefore be isolated with 2 extra residues at its N terminus (glycine and serine) and six extra residues at its C terminus (glutamate to aspartate above). This was not expected to affect any of the functions of the enzyme in any way and it may have had a beneficial effect in helping to increase the solubility of the protein. A fusion protease has been produced with extra amino acids at the N (Pro-Gln-Ile-Thr-) and C termini (-Val-Lys-Gln-Trp) and this did not appear to affect the activity of the protease.[404a]

Transfection of the pGEX-HPR plasmid into *E. coli* strain JM 101 was achieved using the method of Hanahan[416] or a method using rubidium salts (the MOPS method). The Hanahan method uses a complex mixture of metal ions and DMSO (see Experimental section) to make the bacteria competent to take up exogenous DNA. The mechanism of DNA uptake into *E.coli* is unclear, but seems not to involve passive diffusion through pores in the membrane, as no sharp cutoff in transformation probability versus size is seen (the probability of transformation decreases linearly with size of plasmid).[416] The Hanahan method was found to be unreliable, though when it worked well the levels of transformation found were very

high (around 1 x 10^7 transformants μg^{-1} DNA). Therefore, it was abandoned in favour of the more reliable MOPS method which gave around 5 x 10^5 transformants μg^{-1} DNA.

2.6.3 Isolation and analysis of cloning products

Once transformed bacteria were identified by their ampicillin resistance it was necessary to determine which of them carried the recombinant plasmid, which should have been most of them, for the reasons outlined above. Colonies were picked from ampicillin LM agar plates and grown up in 10 ml LB broth containing 100 $\mu g\ ml^{-1}$ ampicillin, induced and harvested after 3 hours. Total cell protein was examined by running PAGE gels on soluble protein from bacteria lysed by boiling in running buffer for 5 mins. Those bacteria that gave a band of the correct mass (37 kDa) were picked, as they were expressing the protease-GST fusion (these were designated JM 101/HPR). The recombinant protein was also identified by comparison to induced bacteria carrying the parent pGEX-2T plasmid, uninduced recombinant bacteria and JM 101. The recombinant bacteria gave a very strong band at 37 kDa that is heavier than the protein (26 kDa) from bacteria carrying only pGEX-2T (JM 101/pGEX) and is not contained in JM 101. This is shown in the gel below.



Column	Sample
1	M. W. standards
2	Soluble protein from JM 101/HPR
3	Insoluble material from JM 101/HPR
4	JM 101 total cell protein
5 & 6	Total cell protein from JM 101/HPR
7	Total cell protein from JM 101/ pGEX

Comparison of lanes 2, 5 and 6 with 7 shows the expected increase in mass on forming the fusion protein. This band is absent in the parent bacteria (lane 4) so must have come from the transformed bacteria. The masses of the standards are 76 kDa, 66.25 kDa, 42.7 kDa, 30 kDa and 17.2 kDa.

Various checks were carried out to ensure that the plasmid isolated from the bacteria selected contained the protease gene. The simplest approach is to perform a comparative restriction analysis on the recombinant and parent plasmids. In this approach the plasmids are digested with two restriction endonucleases that each have one unique site in the plasmid and the size of the fragments produced are compared by agarose gel electrophoresis. A partial restriction map for the two plasmids is given in Appendix B. It can be seen that digestion of the plasmids with Sma 1 and Eco RV will give two bands from both plasmids. The recombinant plasmid will give one of these bands about 300 bp heavier due to incorporation of the HIV-1 protease gene (297 bp). These two enzymes were chosen as they can be used in the same buffer system without either of them losing much activity. This eliminates the need to perform the digests consecutively.

Sequencing of the insert will ensure that the protease is in the correct reading frame and that it is inserted in the correct place in the MCS. Isolation and purification of the plasmid for sequencing proved to be problematic, with the usual protocol giving low yields of the plasmid, heavily contaminated with chromosomal DNA. Small scale preparations, using the Stratagene Plasmid Quik™ kit, were also performed and also gave heavily contaminated plasmid. A modification of the described purification was tried, using benzoylated napthoylated DEAE cellulose (BND cellulose) to purify the plasmid DNA from the chromosomal DNA, tRNA and residual bacterial protein. This also proved problematic, due to the very low yields of

plasmid.

The reason for the low yield is not clear as the parent plasmid (pGEX-2T) is a high copy number plasmid,[408] so pGEX-HPR should also have a high copy number (plasmid copy number is determined by a DNA sequence carried on the plasmid itself). In order to overcome this difficulty the amount of plasmid was amplified by chloramphenicol treatment. When this is added to the bacterial culture host protein synthesis stops but plasmid replication continues as it is dependent on proteins that turn over slowly *e.g.* DNA polymerases I and III. However, chromosomal DNA replication requires newly synthesised proteins, so it is also halted by chloramphenicol treatment. Thus, the amount of plasmid is amplified over chromosomal DNA and host protein by chloramphenicol treatment, which makes the purification much easier. This strategy was found to be effective, increasing the yield of plasmid around 50-fold (to around 300-500 μg per litre of culture).

It was possible that a reason for the low yield of plasmid was that after a few hours culturing the ampicillin in the medium had been degraded by ß-lactamases secreted into the medium by the resistant bacteria. After this point any bacterium that spontaneously lost the plasmid would be able to grow, and grow much more quickly than those bacteria that still retained the plasmid. These non-transformed bacteria would quickly dominate the bacterial population and result in very low yields of plasmid. To check this hypothesis a viable cell count was performed. A fresh culture was plated out on LB plates containing ampicillin (100 μg ml^{-1}), the rest of the culture was incubated overnight and then plated out as before. The number and size of the colonies on the plates after 24 hours growth were assessed and found to be equivalent. This indicates that the plasmid is well maintained over long growth periods. The low yields of plasmid in the absence of chloramphenicol amplification therefore remains unexplained.

When plasmid preparations were performed to isolate further amounts of pGEX-2T for cloning the purity of the plasmid was checked by agarose gel electrophoresis on the intact plasmid and on the plasmid restricted with Sma 1 and Eco RV. These were run against similarly treated samples of the pure plasmid.

Once isolated and purified attempts were made to sequence the plasmid DNA. Unfortunately, sequencing proved extremely difficult. The dideoxy method,[417] was used, with ^{35}S as the radiolabel. There seemed to be impurities in the plasmid preparation that interfered with the DNA polymerase used in sequencing and so at best only partial sequences of the pGEX-HPR around the MCS were obtained. An

alternative explanation is that the denaturation protocol of the double strand plasmid DNA (heating in the presence of sodium hydroxide) used was not working well.

In an effort to circumvent the problems encountered in sequencing double stranded plasmid DNA, single strand sequencing was examined instead as sequencing single strand DNA is usually much easier. Two alternative approaches to obtaining single strand DNA were investigated.

i) Clone the protease gene into the bacteriophage M13. M13 produces single strand DNA when it infects bacteria, such as JM 101, that express the appropriate cell surface receptor.

ii) Make pGEX-HPR into a phagemid by cloning the origin of replication (ori) from the bacteriophage M13 into it. This will make a bacterial cell transfected with the phagemid produce large quantities of single strand DNA. The ori will be amplified by PCR from a suitable M13 vector.

Approach (i) was taken first. Cloning into M13 was carried out in a very similar way to the cloning into pGEX-2T discussed above. The M13mp18 vector[418] was used as the double stranded replicative form (RF). This was handled exactly as if it were a plasmid. Like pGEX-2T M13mp18 RF has a multiple cloning site (MCS) containing, amongst others, unique Bam H1 and Eco R1 sites. The MCS is located within two genes required for lactose metabolism. Thus, following the same procedure as for the cloning of the protease, the PCR amplified protease gene was ligated into the MCS of M13mp18. The recombinant M13 vector was then transfected into competent JM 101 in the same way as pGEX-HPR was.

Care was taken to ensure that the competent bacteria used in the transfection retained the receptor for M13, the sex pili, which is encoded by a bacterial plasmid known as the F' episome. This episome also contains two genes (*pro* A and B) involved in proline biosynthesis. JM 101 does not contain these genes, so growing JM 101 on a minimal medium, such as M9, that does not contain proline will ensure that the F' episome is retained. Thus, the bacteria that were made competent for this procedure were grown from colonies picked from an M9 minimal agar plate.

Transformation was carried out exactly as above and the cells plated out in semi-solid medium (top agar) on LB agar plates. The screening was carried out by identifying clear plaques on a bacterial lawn from the transformation. M13 slows the growth of infected bacteria, resulting in areas of low bacterial growth that show up as clear plaques against the lawn of quickly growing, uninfected bacteria. Metabolism of

BCIG (X-gal) in the top agar by lactose metabolising bacteria will result in a blue by-product, which makes the bacterial plaque blue. JM 101 is unable to metabolise lactose, having had two of the genes required deleted. However, these two genes are carried on the F' episome. Thus, those plaques that have a blue colour arise from JM 101 infected with M13 that contains functional lactose metabolising genes. Successful cloning of a gene into the M13 MCS will result in insertional inactivation of the two lactose metabolising genes it carries, which means no metabolism of the BCIG to the blue compound. Thus, clear plaques arise from bacteria infected with M13 that contains the gene to be cloned. Once clear plaques were identified single strand DNA was prepared from them and sequenced. Similar problems were encountered in sequencing this DNA as were found for sequencing double strand plasmid DNA. Only a partial sequence was obtained by this method (see Figure 2.6.4).

The phagemid approach was explored initially but several problems were encountered early on and so the route was set aside in favour of the M13 procedure. The PCR amplified origin of replication (ori) from M13 was cloned into the Aat II site of pGEX-2T. The recombinant bacteria harbouring the plasmids required were identified after plating transformed bacteria onto agar plates containing ampicillin. Since the recombinant plasmid contains the M13 ori it is slightly larger than the non-recombinant plasmid. The difference in the size of the intact plasmids is too small to reliably detect by gel electrophoresis (the M13 ori is 123 b.p. and pGEX-2T is 4948 b.p.). However, restriction analysis of plasmid isolated from ampicillin resistant bacteria should have allowed those plasmids containing the ori to be easily identified. Several possible restriction digests were considered, but none gave reliable results and so the route was abandoned. This may be due to the method used to isolate the plasmid, STET preparations, instead of the much more time-consuming alkaline lysis/cesium chloride banding procedure, which gives much purer plasmid, used for large-scale preparations. The impurities in the plasmid isolated by the STET procedure may interfere with the action of the restriction enzymes and result in poor reproducibility of the digests.

Unfortunately, only partial sequences were ever obtained from the above efforts at sequencing. The most extensive sequence obtained gave the 3' 96 nucleotides of the protease gene but this did not show the site of insertion as it was obtained using the 3' PCR primer.

5'...ATA/GAA/TCT/GTG/GAC/ATA/AAG/CTA/TAG/TAG/GAC/CTA/CAC/CTT/GTC/AA
C/ATA/ATT/GGA/AGA/AAT/CTG/TTG/ACTCAG/ATT/GGT/TGC/ACT/TTA/AAT/TTT 3'

Figure 2.6.4: Partial sequence of HIV-1 protease cloned into pGEX-2T

2.6.4 Expression and purification strategy

Once the fusion protein had been expressed in the recombinant *E. coli* it was thought that it could be very easily be purified by affinity chromatography. Simply passing the crude cell extract over a glutathione agarose column should result in extensive purification of the HIV protease-GST fusion protein. The active site of glutathione-S-transferase (GST) clearly has a high affinity for the glutathione and will bind to it and the remainder of the cellular proteins will pass through the column. Elution with a glutathione solution releases substantially pure fusion protein. Treatment with thrombin will then release the protease, which can then be further purified by affinity chromatography over a pepstatin agarose column, as has been used before.[419] Pepstatin is a generalised aspartic protease inhibitor, and so will be bound specifically by the protease. Any residual cellular proteins and the GST will pass through unhindered and the protease will be retained. Elution should give essentially homogenous protein.

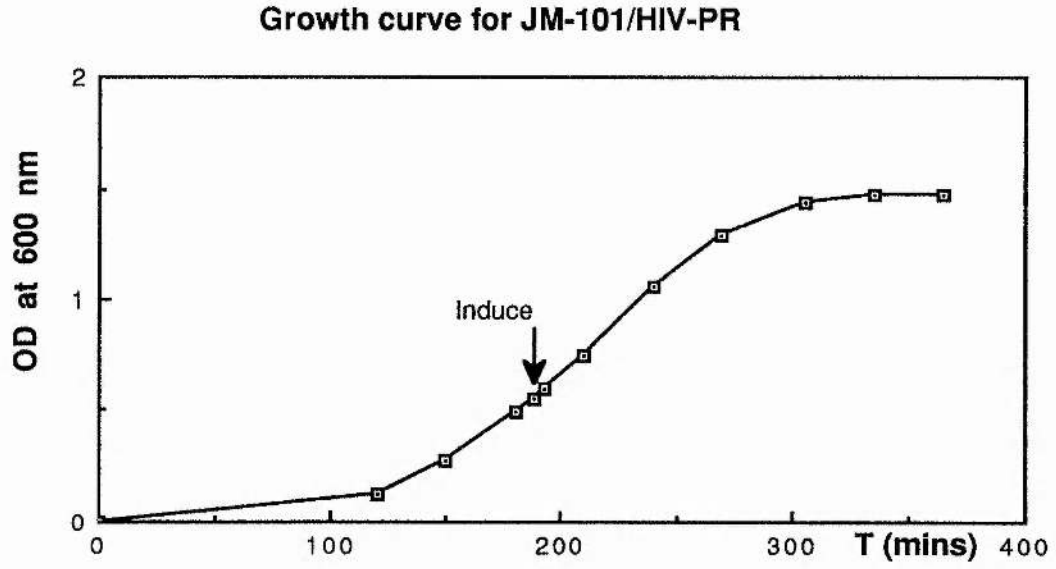
An inherent problem with the use of a fusion protein approach for expression and purification of the HIV protease is that the protease may not be active as a fusion. The enzyme is an obligate dimer (see Section 1.10.2.3), with the dimerisation interface lying at the N and C-termini. The N-terminus of the protease will be occluded by the presumably folded GST, thus preventing dimerisation. Prevention of dimerisation will prevent expression of activity. This means that the purification cannot be followed in the usual way, by assessing protease activity in fractions collected from columns and so forth. Purification therefore has to be followed by gel electrophoresis. Most attempts to express the protease as a fusion have resulted in active protein[404i,k] while others have not.[419] In this case activity could only be restored by denaturation in urea and refolding. It would be expected that the protease would be active as a fusion protein as it is expressed *in vivo* as a fusion with the rest of the *gag-pol* polyprotein. Kotler *et al.*[420] have found that the protease

is active when its N-terminus is fused with parts of the *gag* polyprotein. Thus, the factors that determine the activity of a protease fusion are manifold and difficult to determine.

During the initial stages of the purification the distribution of the fusion protein between the supernatant from sonication and the cell debris was checked by SDS-PAGE. It was found that some of the fusion remained in the cell debris, *i.e.* it was produced as an insoluble aggregate. In an attempt to maximise the level of soluble protein a series of experiments on the effect of different induction conditions were undertaken.

Lower concentrations of IPTG will reduce the rate of production of the fusion protein and thus reduce the chances for aggregation. Concentrations from 20-150 μM IPTG were tried and the proportion of the insoluble protein did not seem to alter greatly over this range. Hartmann *et al.*[421] have reported that low growth temperatures also result in reduced amounts of recombinant proteins being formed as inclusion bodies. However, downshifting the growth temperature to 25 °C did not reduce the amount of insoluble protein. Lastly, the effect of induction time (from 1-5 hours) was examined for its effect both on the proportion of insoluble protein, the level of expression and the cell density.

To assess the effect of the expression of the protease on the growth rate of JM-101 the optical density at 600 nm of the culture was monitored. The results are presented graphically below. It can easily be seen that the expression of the protease does not cause the saturation density of the culture to be lowered significantly *i.e.* the expression of protease is not toxic to the cell. This is unusual, as most groups have observed extremely deleterious effects on cell growth occurring on induction of protease expression.[404] Saturation density is reached approximately 2 hours post-induction. However, since no activity could be obtained in crude preparations the effect of induction time on the level of activity produced could not be assessed.



To select the optimal time post-induction for harvesting the cells the level of fusion protein was monitored at various times (30 mins.-3.5 hours) by denaturing PAGE. The levels in both total cell protein and in supernatants of lysozyme lysates of 1.5 ml aliquots of the culture were examined. The level of protein expressed was estimated from the intensity of the band corresponding to the fusion protein when the gels were loaded under identical conditions. Maximal expression was achieved 2-2.5 hours post-induction.

Once optimal growth conditions had been established the transformed cells were grown, harvested, sonicated and centrifuged to remove insoluble cell debris. The supernatant, containing most of the HIV PR-GST fusion, was applied to a glutathione agarose column and the wash-through collected. The GST fusion was eluted by washing with 2 column volumes of 5 mM glutathione. However, no protein was eluted from the column.

When the effluent from column loading and the column washing were run on denaturing polyacrylamide gels there was a strong band from both sets of fractions of a molecular weight around 29 kDa. When the total cell protein was run on PAGE gels the major band was observed, as expected, at 37 kDa. Thus there is a loss of around 8 kDa from the recombinant fusion protein on sonication. This implies that the fusion protein was being cleaved on sonication to give the smaller form observed, and this form was not binding to the glutathione column. If the fusion protein were truncated before expression *i.e.* at the gene level, then the gene has been cleaved before

expression. This is unlikely as the host strain JM 101 contains only one restriction endonuclease, Eco K, which has an 8 b.p. cleavage site. This sequence is not contained in the GST or the protease genes. Also, if the plasmid were to be cleaved in JM 101 then it would not be maintained from generation to generation in the JM 101, thus there would be no source of ampicillin resistance and the bacteria would die in culture.

The cleavage therefore takes place at the protein level and could be due to either cellular protease(s) or the HIV-1 protease. As the cleavage takes place over a very short period of time and only after sonication it is unlikely that a cellular protease is the cause. Also, expression of non-proteolytic enzymes as GST fusions in JM101 does not result in any detectable cleavage of the fusion,[412] again supporting the fused protease as the source of the cleavage. Since cleavage is not seen in the total cell protein, sonication must somehow activate the protease, perhaps by breaking up aggregations of the fusion protein in solution (or perhaps even solid inclusion bodies).

The cleavage was almost definitely not occurring in the protease, as autoproteolysis of the protease does not release a fragment of the correct size,[388] and autoproteolysis occurs on a much slower time-scale than that observed here. The protease therefore cleaves the GST. The observed loss in mass on cleavage of the fusion protein (around 8 kDa) is consistent with cleavage at a putative good protease site in GST, Phe 53-Pro 54, which results in the loss of 7.5 kDa. It has been shown that the main determinant of cleavage by the protease is not the sequence but the conformation of residues in the vicinity of the scissile bond (see Section 1.8.3). Thus, the possibility of the Phe 53-Pro 54 site being cleaved was investigated by examining the predicted secondary structure of the 20 amino acid region around the putative cleavage site with that around known Phe-Pro cleavage sites from the *gag-pol* polyprotein. Secondary structure prediction was undertaken using the GeneWorks™ program and the PepPlot and Peptide Structures routines from the UWGCG package.[422] No definite secondary structure prediction was made which is consistent with studies on peptide substrates for the HIV protease that show they have only random conformation in solution.[372]

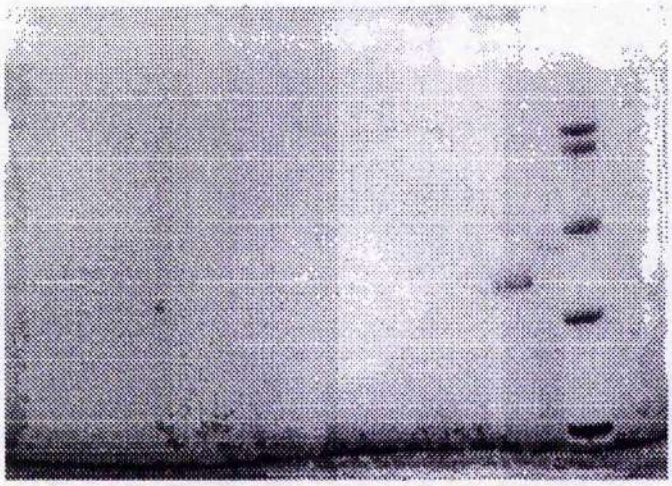
The recent solution of the crystal structure of GST from rat[423] shows that the Phe-Pro residues are in the glutathione-binding domain. The equivalent residues to Phe-53 and Pro-54 in the rat GST are contained in a *cis*-Pro bend, which is on the exterior of the protein. This means that the putative cleavage site is accessible to the protease, which is consistent with Phe-53-Pro-54 being the actual site of cleavage. This could be demonstrated by isolating the small fragment and sequencing its N-terminus.

In an effort to suppress the cleavage by inhibiting the break-up of the putative protein aggregates sonications were carried out on cells suspended in decreasing volumes of sonication buffer. This also mimics conditions in the cell as closely as possible, which would be helpful if some cellular factor were either inhibiting the protease or promoting its aggregation. This approach was not fruitful, which is not unexpected as the cleavage is an intermolecular reaction, so would be expected to proceed faster in higher concentrations. Sonications were also tried using very large volumes of buffer to reduce the rate of this intermolecular reaction. This proved ineffective. The use of raised pH (from 7.5 to 9 in 0.5 unit steps) to reduce the activity of the protease (which has a pH optimum around 5.5, see Section 1.8.3) was also ineffective.

Sonication in the presence of 20 μM pepstatin, which is a general inhibitor of aspartic proteases including the HIV protease (K_i = 1.4 μM[362]), should have prevented this cleavage. This was not the case, however, and so the source of the cleavage remained unclear. It may be that the reaction occurs too rapidly for the pepstatin to interact with the protease and inhibit it, perhaps because pepstatin has a two-stage interaction with the HIV protease as it does with other aspartic proteases.[424] The second stage of the interaction occurs over the course of several minutes, so if pepstatin interacts with the protease in this way then a high level of inhibition will only be achieved after a considerable period of time. This delay in achieving good inhibition will mean that the cleavage reaction will have finished before the enzyme is inhibited. In order to overcome this it was decided to incubate the cells in the presence of 50 μM pepstatin for 20 minutes before sonication in the presence of 20 μM pepstatin. This also had no observable effect on the cleavage. Thus, it appears that while the cleavage is not caused by a cellular protease, inhibiting the HIV protease is problematic.

The difficulties discussed above meant that no uncleaved protease fusion was isolated from the recombinant bacteria. Attempts were made to purify the cleaved

fragment and thrombinise off the cleaved GST after purification. Precipitation of the cleaved fusion was efficiently achieved with ammonium sulfate at 20% saturation, with few contaminating proteins (purity was estimated to be around 80%). There was still no measurable protease activity. Subsequent size exclusion chromatography on Sephadex G-75 also gave no activity. The purity of the protein is shown in the PAGE gel below.



The rightmost lane contains the molecular weight markers, the adjacent lane shows a sample of the protease after elution from the G-75 Sephadex column. The masses of the standards are 76 kDa, 66.25 kDa, 42.7 kDa, 30 kDa and 17.2 kDa.

It has recently been brought to our attention that similar attempts in other groups to express the HIV-1 protease as a GST fusion have all failed, with no or very low amounts active protein isolated.[425] The inconsistency between the lack of protease activity against the peptide substrate used in the *in vitro* assays and the proposed high level of activity against GST remains unexplained.

2.6.5 Future Work

A solution to the problem of cleavage is to use site-directed mutagenesis to mutate away the cleavage site in the GST. Since the proline at position 54 causes a bend in the protein chain, mutation of this residue is likely to cause gross changes in the folding of the GST, probably resulting in a protein that will not bind to glutathione. Thus, it was decided to mutate the phenylalanine to glutamine as this residue is of

approximately the same size as the phenylalanine side-chain, so will not perturb the side-chain packing greatly. It is also known that the Gln-Pro linkage is not cleaved by the HIV-1 protease.[379] The protease may also be expressed as a fusion with a different protein. In order to observe activity it may be necessary to express the protease as an insoluble fusion protein (in inclusion bodies) and return activity by refolding solubilised inclusion bodies from urea.

3.0 Experimental

Optical rotations were measured at room temperature on Optical Activity AA-100 and AA-1000 polarimeters.

Elemental analyses were performed in the departmental microanalytical laboratory and by the University College, London laboratory.

N.m.r. spectra were taken on Bruker AM-360 (360 MHz, f.t. ^{1}H-n.m.r. and 89.71 MHz ^{13}C-n.m.r.), Bruker AM-300 (300 MHz, f.t. ^{1}H-n.m.r., 121.49 MHz, ^{31}P-n.m.r. and 74.76 MHz, ^{13}C-n.m.r.), Jeol JNM-GX270 (270 MHz, f.t., ^{1}H-n.m.r.), Varian Gemini 200 (200 MHz, f.t. ^{1}H-n.m.r. and 50.31 MHz,^{13}C-n.m.r.) and Jeol FX-90Q (90 MHz, f.t. ^{1}H-n.m.r.) spectrometers. ^{1}H-N.m.r. spectra are described in parts per million downfield shift from TMS and are reported consecutively as position (δ_H), relative integral, multiplicity (s.-singlet, d.-doublet, t.-triplet, q.-quartet, m.-multiplet, d.d.-double doublet and br.-broad), coupling constant and assignment (numbering according to the IUPAC nomenclature for that compound[426]). For compounds run in C^2HCl_3, the spectra were referenced to the internal $CHCl_3$ signal at 7.27 p.p.m. and for compounds run in ^2H_2O the spectra were referenced to the internal HO^2H peak at 4.60 p.p.m., or dioxan at 3.66 p.p.m.

{^{1}H}^{13}C-Nuclear magnetic resonance spectra were referenced to chloroform (at 77.20 p.p.m.), dioxan (67.40 p.p.m.) or dimethyl sulfoxide (at 39.70 p.p.m.). and are reported consecutively as position (δ_H), multiplicity (d.-doublet), coupling constant and assignment (quat.- quat.).

Conformational effects are indicated by the symbol † appearing after the chemical shift of the atom(s) concerned.

I.r. spectra were taken on Perkin-Elmer Series 1500 and Perkin-Elmer 1710 f.t. i.r. spectrometers. The samples were prepared as Nujol mulls, solutions in chloroform or thin films between sodium chloride discs. Absorption maxima are given in wavenumbers (cm^{-1}) relative to a polystyrene standard.

Mass spectra and accurate mass measurements were recorded on a VG 70-250 SE, a Kratos MS-50 or by the SERC service at Swansea using a VG AZB-E. Fast atom bombardment spectra were recorded using glycerol as a matrix. Major fragments are

given as percentages of the base peak density (100%).
UV/ Vis spectra were recorded on Pye-Unicam SP8-500 or SP8-100 spectrometers. Flash chromatography was performed according to the method of Still[427] using Sorbsil C 60 (40-60 μm mesh) or Machery-Nagel silica gel N.

Analytical thin layer chromatography was carried out on 0.25 mm precoated silica gel plates (Macherey-Nagel SIL G/UV_{254}) and compounds were visualised using UV fluorescence, iodine vapour, aqueous potassium permanagnate or ninhydrin.
Melting points were taken on an Electrothermal melting point apparatus and are uncorrected.
All solvents were distilled before use and light petroleum refers to that portion boiling between 40 and 60°C. Solvents were dried according to the literature procedures.[428] Agarose gel electrophoresis was carried out on 1% agarose gels containing 1 μg ml^{-1} ethidium bromide with TBE buffer at 100 mA unless otherwise specified. Sonications were carried out on a MSE Soniprep 150 sonicator at full power. Centrifugations were carried out on a Beckman JA-2000 centrifuge in JA-14 or JA-6 rotors and ultracentrifugations on a Beckman TL-100 in a J-2 rotor. Oligonucleotides were synthesised on an Applied Biosystems 381A DNA synthesiser. DNA concentrations were measured by absorbance at 260 nm. Common broths, buffers and stock solutions were made as described by Sanbrook *et al.*[429]
Molecular modelling calculations were carried out on a Silicon Graphics IRIS 4D/85GT and an Indigo Elan R4000. Macromodel V.3.1X[430] and INSIGHT 2.1.2[431] were used to perform these calculations and display the results. Crystal structures were obtained from the Protein Data Bank.[432]

N-*tert*-Butoxycarbonyl-(2S)-phenylalanine phenylhydrazide (1).-

To a stirred solution of N-*tert*-butoxycarbonyl-(2S)-phenyalanine (265 mg, 1 mmol) in dry DMF (10 ml) at 0 °C was added N-methylmorpholine (109 μl, 1 mmol), followed by 2-methylpropylchloroformate (136 μl, 1 mmol). After 1.5 h. phenylhydrazine (98.3 μl, 1 mmol) and N-methylmorpholine (109 μl, 1 mmol) were added in DMF (10 ml) and the mixture stored at 4 °C overnight. The solution was then partioned between ethyl acetate (40 ml) and water (10 ml), the ethyl acetate

layer was separated and washed with 1 M HCl (3 x 10 ml), water (10 ml), 5% sodium bicarbonate solution (3 x 10 ml) and water (10 ml). The organic layer was dried ($MgSO_4$) and the solvent removed under reduced pressure to give a yellow oil which was crystallised from methanol/ water to give cream crystals (173 mg, 48%) of m.p 135-137 °C. $[\alpha]_D^{22}$ -16.1° (c 1 in MeOH); ν_{max} ($CHCl_3$) 3429 (NH), 3300 (amide NH), 1750-1650 (CO) and 756 cm^{-1} (aromatic); λ_{max} (MeOH) 283 nm (ε 2060); δ_H (90 MHz, C^2HCl_3) 1.40 (9H, s., $(CH_3)_3C$), 3.03 (2H, d., *J* 8.1 Hz, CH_2), 4.48 (1H, m., CH), 5.26 (1H, d., *J* 6.9 Hz, PhNH), 6.58 (1H, d., *J* 6.9 Hz, amide NH), 7.1-7.4 (10 H, m., aromatic) and 8.2 (1H, br.s., urethane NH); *m/z* (CI) 356 ($[M+H]^+$, 5.1%), 299 ($[M+H-(CH_3)_3C]^+$, 25.9), 282 ($[M-(CH_3)_3CO]^+$, 68.4), 256 ($[M-(CH_3)_3COC(O)]^+$, 100) and 120 ($NH_2CH(NH)CH_2Ph^+$, 98.1).

N-*tert*-Butoxycarbonyl-(2S)-phenylalanine nitrophenylhydrazide (2).-

This was prepared in an identical manner to the phenylhydrazide (**1**), using nitrophenylhydrazine (153 mg, 1 mmol). The product was obtained as beige amorphous crystals, after recrystallisation from methanol/ water (192 mg, 48%) of m.p. 208-209 °C (dec.). $[\alpha]_D^{22}$ -8.9° (c 1 in MeOH); ν_{max} (Nujol) 3340-3250 (NH), 1689 (urethane CO), 1653 (amide CO), 1531 (NO_2), 1329 (NO_2), 843 (*p*-substituted aromatic) and 724 cm^{-1} (mono-substituted aromatic); λ_{max} (MeOH) 353 nm (ε 56,800); δ_H (360 MHz, C^2HCl_3) 1.46 (9H, s., $(CH_3)_3C$), 3.06 (2H, m., CH_2), 4.36 (1H, m., CH), 6.65 (2H, d., *J* 8.9 Hz, nitroaromatic), 7.42 (5H, s., Ph), 8.04 (2H, d., *J* 8.9 Hz, nitroaromatic), 9.20 (1H, s., amide NH) and 10.37 (1H, s., urethane NH); *m/z* (FAB) 401 ($[M+H]^+$, 12.3%), 300 ($[M+H - (CH_3)_3COC(O)]^+$, 30.7), 91 ($PhCH_2^+$, 13.6) and 57 ($(CH_3)_3C^+$, 100).

Benzylhydrazine (3).-

Benzylhydrazine dihydrochloride (1.37 g, 7 mmol) was added to a stirred solution of 1 M sodium hydroxide solution (15 ml) at room temperature. After 1 h. and the aqueous phase was extracted with chloroform (3 x 20 ml). The extracts were combined, dried ($MgSO_4$), filtered and the solvent removed under reduced pressure to give a yellow oil (702 mg, 82%). ν_{max} ($CHCl_3$) 3018 (NH) and 754 cm^{-1} (aromatic);

δ_H (90 MHz, C^2HCl_3) 2.81 (2H, s., CH_2Ph) and 7.31 (5H, s., aromatic).

N-*tert*-Butoxycarbonyl-(2S)-phenylalanine benzylhydrazide (4).-

This was prepared as for the phenyl analogue (**1**), using benzylhydrazine (122 mg, 1 mmol). Recrystallisation from methanol/ diethyl ether gave the product as a white solid (122 mg, 49%) of m.p. 121-123 °C. $[\alpha]_D^{22}$ -7.2° (c 2 in $CHCl_3$); ν_{max} ($CHCl_3$) 2981 (amide NH), 1705 (urethane CO), 1680 (amide CO) and 710-750 cm^{-1} (aromatic); λ_{max} (MeOH) 283 nm (ε 2240); δ_H (270 MHz, C^2HCl_3) 1.40 (9H, s., $(CH_3)_3C$), 3.04 (2H, d., *J* 7.5 Hz, $C\underline{H}_2NH$), 3.83 (2H, m., $CHC\underline{H}_2$), 4.27 (1H, m., C-2 H) and 7.06-7.39 (10H, m., aromatics); *m/z* (CI) 370 ([*M*+H]$^+$, 4.2%), 314 ([*M*+H-$(CH_3)_3C$]$^+$, 25.4), 296 ([*M* -$((CH_3)_3C+H_2O)$]$^+$, 100), 270 ([*M* -$(CH_3)_3CCO_2$]$^+$, 49.1) and 108 ($PhCH_2NH_3^+$, 17.3).

Methyl (2S)-tyrosinate hydrochloride (5).-[433]

(2S)-Tyrosine (1 g, 5.5 mmol) was dissolved in dry methanol (15 ml) at 0 °C and distilled thionyl chloride (1.69 g, 14 mmol) was added dropwise with stirring, keeping the temperature at 0 °C. The mixture was then refluxed for 90 min. The solvent was removed under reduced pressure to give an off-white solid, which was recrystallised from methanol/ diethyl ether to give the product as white crystals (1.21 g, 95%) of m.p 189-191 °C (lit.[433] 190 °C). *m/z* (Found: (*M* + H$^+$) 196.0973. $C_{10}H_{14}NO_3$ requires 196.0973); $[\alpha]_D^{22}$ +65.2° (c 2.4 in H_2O) (lit.[433] $[\alpha]_D^{22}$ +74° (c 3 in pyridine)); ν_{max} (Nujol) 3338 (NH), 1743 (CO) and 750 cm^{-1} (aromatic); δ_H (90 MHz, ^2H_2O) 2.97 (2H, d., *J* 6.5 Hz, CH_2), 3.64 (3H, s., OCH_3), 4.15 (1H, t., C-2 H), 6.65 (2H, d., *J* 8.2 Hz, aromatic), 6.92 (2H, d., *J* 8.2 Hz, aromatic); δ_C (50.31 MHz, ^2H_2O) 37.56 (CH_2Ph), 56.34 ($CHNH_3^+$), 56.99 (CH_3O), 118.77 (*ortho* aromatic), 128.18 (*para* aromatic), 133.64 (*meta* aromatic), 158.03 (*ipso* aromatic) and 172.93 (ester CO); *m/z* (EI) 195 ([*M*-HCl]$^+$, 6.5%), 136 ([*M*-HCl-CO_2Me]$^+$, 33.9), 107 ($C_7H_6OH^+$, 100) and 77 ($C_6H_5^+$, 16.4).

(2S)-Tyrosine methyl ester (6).-

(2S)-Tyrosine methyl ester hydrochloride (695 mg, 3 mmol) was dissolved in saturated sodium bicarbonate solution (20 ml). The solution was then extracted with chloroform (4 x 15 ml), the organic layer was dried (Na_2SO_4) and the solvent removed under reduced pressure to give a white solid (421 mg, 72%) of m.p. 132-134 °C (lit.[434] 133-135 °C). $[\alpha]_D^{22}$ +24° (c 4 in MeOH); ν_{max} ($CHCl_3$) 3353 (NH), 2850-2640 (aromatic) and 1743 cm^{-1} (CO); δ_H (90 MHz, C^2HCl_3) 2.94 (2H, app. d., *J* 6.8 Hz, CH_2), 3.65 (3H, s., OCH_3), 4.13 (1H, m., C-2 H), 6.65 (2H, d., *J* 7.9 Hz, aromatic) and 6.95 (2H, d., *J* 7.9 Hz, aromatic); *m/z* (EI) 195 (M^+, 4%), 136 ([*M*-CO_2Me]$^+$, 16), 107 ($C_7H_7OH^+$, 42) and 77 ($C_6H_5^+$, 100).

Methyl N-*tert*- butoxycarbonyl-(2S)-phenylalanyl-(2S)-tyrosinate (7).-

N-*tert*-Butoxycarbonyl-(2S)-phenylalanine (133 mg, 0.5 mmol) and N-methylmorpholine (55 µl, 0.5 mmol) were dissolved in dry DMF (10 ml) and 2-methylpropylchloroformate (65 µl, 0.5 mmol) added. The reaction was stirred at 0 °C for 1.5 h., methyl ester (**6**) (98 mg, 0.5 mmol) added in dry DMF (5 ml) and the reaction stirred overnight at 4 °C. Water (15 ml) was added and the mixture extracted into ethyl acetate (40 ml). The ethyl acetate layer was separated and washed with 1 M HCl (3 x 10 ml), water (10 ml), saturated sodium bicarbonate (3 x 10 ml) and water (10 ml), dried ($MgSO_4$) and the solvent removed under reduced pressure to give a yellow oil. This was twice recrystallised from ethyl acetate/ light petroleum to give a white solid (73 mg, 15.8%) of m.p. 142-143 °C. ν_{max} ($CHCl_3$) 1748 (ester CO) cm^{-1}; δ_H (90 MHz, C^2HCl_3) 1.40 (9H, s., $(CH_3)_3C$), 3.0 (4H, m., CH_2 x2), 3.68 (3H, s., OCH_3), 4.34 (1H, m., Phe CH), 4.76 (1H, m., Tyr CH), 5.27 (1H, d., *J* 7.8 Hz, amide NH), 6.66 (2H, d., *J* 7.5 Hz, tyrosine aromatic), 6.87 (2H, d., *J* 7.5 Hz, tyrosine aromatic), 7.26 (5H, s., Ph) and 8.26 (1H, br., urethane NH).

(2S)-Tyrosine amide (8).-

Methyl ester hydrochloride (**5**) (0.8 g, 3.25 mmol) was dissolved in saturated methanolic ammonia solution (15 ml) and stirred at room temperature for 96 h. The ammonia was blown off in a stream of nitrogen and the solvent removed under reduced pressure to give a cream solid, which was recrystallised from ethanol/ light petroleum to give white crystals (0.52 g, 69%) of m.p. 240-243 °C. *m/z* (Found: $(M+H)^+$ 181.0977. $C_9H_{13}N_2O_2$ requires 181.0977); $[\alpha]_D^{22}$ +21° (c 1 in H_2O); ν_{max} (Nujol) 3330 (amide NH), 3175 (amine NH), 1703 (CO), 1458 (amide II) and 793 cm^{-1} (aromatic); δ_H (90 MHz, ^2H_2O) 3.12 (1H, d.d., J_{AX} 7.4 Hz, J_{AB} 13.8 Hz, 1 of CH_2), 3.21 (1H, d.d., J_{BX} 5.8 Hz, J_{AB} 13.8 Hz, 1 of CH_2), 4.22 (1H, d.d., J_{AX} 7.5 Hz, J_{BX} 5.8 Hz, CH), 6.91 (2H, d., J 8.8 Hz, aromatic) and 7.23 (2H, d., J 8.8 Hz, aromatic); δ_C (50.31 MHz, d_6-DMSO) 36.23 (CH_2Ph), 53.92 (CH-NH), 115.53 (*ortho* aromatic), 125.26 (*para* aromatic), 130.74 (*meta* aromatic), 156.79 (*ipso* aromatic) and 170.16 (amide CO); *m/z* (EI) 180 (M^+, 2.1%), 163 ($[M-OH]^+$, 58.5), 136 ($[M-CONH_2]^+$, 95.5), 107 ($CH_2C_6H_4OH^+$, 100) and 74 ($NH_3CONH_2^+$, 72.6).

N-*tert*-Butoxycarbonyl-(2S)-leucine hydrate (9).-

To a stirred solution of (2S)-leucine (1.12 g, 10 mmol) in 1 M sodium hydroxide (10 ml) and distilled *iso*propanol (10 ml) was added di-tertiary butyl dicarbonate (2.9 g, 13 mmol) in *iso*propanol (5 ml). The resulting precipitate was dissolved in 1 M sodium hydroxide (7 ml). After 1 h., the alcoholic layer was removed under reduced pressure, the aqueous layer washed with light petroleum (15 ml) and acidified to pH 3 with 1 M sulphuric acid. The resulting solid was filtered and carefully dessicated over phosphorous pentoxide to give the product as white crystals (1.66 g, 81%) of m.p 68-71 °C (lit.[435] 67-71 °C). *m/z* (Found: $(M+H)^+$ 232.1549. $C_{11}H_{22}NO_4$ requires 232.1549); $[\alpha]_D^{20}$ -21° (c 2 in AcOH) (lit.[435] $[\alpha]_D^{20}$ -25° (c 2 in AcOH)); ν_{max} ($CHCl_3$) 3442 (NH), 3019 (CH), 1714 (carboxyl CO) and 1694 cm^{-1} (urethane CO); δ_H (270 MHz, C^2HCl_3) 0.96 (6H, d., J 6.5 Hz, $(C\underline{H}_3)_2CH$), 1.45 (9H, s., $(CH_3)_3C$), 1.47-1.83 (3H, m., $(CH_3)_2C\underline{H}$ and CH_2), 4.36 (1H, m., $C\underline{H}NH$), 4.97 (1H, d., J 7.9 Hz, NH) and 9.22 (1H, br. s., CO_2H); δ_C (50.31 MHz, C^2HCl_3) 22.23 & 23.32 ($CH(\underline{C}H_3)_2$), 25.31 ($\underline{C}H(CH_3)_2$), 28.76 ($(\underline{C}H_3)_3C$), 41.92 (CH_2), 52.61 (CHNH), 80.61 ($(CH_3)_3\underline{C}$), 156.19 (urethane CO) and 178.37 (acid CO); *m/z* (CI) 249

([*M*+NH_4]$^+$, 38%), 232 ([*M*+H]$^+$, 1.8), 186 ([*M* -CO_2H]$^+$, 20.04), 130 ([*M* -$(CH_3)_3COCO$]$^+$, 49.04), 86 ($NHCH(CH_2)CO_2H^+$, 53.4) and 57 ($(CH_3)_3C^+$, 100).

N-*tert*-Butoxycarbonyl-(2S)-leucyl-(2S)-tyrosine amide (10).-

N-*tert*-Butoxycarbonyl-(2S)-leucine hydrate (**9**) (231 mg, 1 mmol) was dissolved in ethyl acetate (10 ml), dried ($MgSO_4$) and the solvent evaporated to give a clear oil. The oil was then dissolved in dry DMF (10 ml), containing N-methyl morpholine (109 µl, 1 mmol) at 0 °C, and 2-methylpropylchloroformate (136 µl, 1 mmol) added with stirring. The reaction mixture stirred for 90 min. at 0 °C. A solution of amide (**8**) (165 mg, 1 mmol) and N-methyl morpholine (109 µl, 1 mmol) in DMF (3 ml) was added dropwise and the reaction stirred overnight at 4 °C. The reaction was partitioned between water (15 ml) and ethyl acetate (3 x 10 ml), which was then washed with 1 M HCl (3 x 10 ml), water (10 ml), 5% sodium bicarbonate solution (3 x 10 ml) and water (10 ml). The organic layer was dried ($MgSO_4$) and the solvent removed under reduced pressure to yield a white solid, which was recrystallised from ethyl acetate/ light petroleum to give the product as white crystals (66 mg; 16.5%) of m.p. 119-122 °C. *m/z* (Found: (*M* + H)$^+$ 394.2342. $C_{20}H_{32}N_3O_5$ requires 394.23417); $[\alpha]_D^{20}$ +18.8° (c 1 in MeOH); ν_{max} (Nujol) 3541 (OH), 3470-92 (NH), 1699 (urethane CO), 1685 & 1669 (amide CO) and 753 cm^{-1} (aromatic); δ_H (270 MHz, d_6-DMSO) 0.89 (6H, m., $(C\underline{H}_3)_2CH$), 1.37 (9H, s., $(CH_3)_3C$), 1.55 (2H, m., $CHC\underline{H}_2$), 1.62 (1H, m., $C\underline{H}CH_2$), 2.57 (1H, d.d., J_{AB} 14.2 Hz, J_{AX} 8.9 Hz, 1 of CH_2Ph), 2.88 (1H, d.d., J_{AB} 14.2 Hz, J_{BX} 5.5 Hz, 1 of CH_2Ph), 3.81 (1H, m., NHCH Tyr), 4.33 (1H, m., NHCH Leu), 6.62 (2H, d., *J* 8 Hz, aromatic), 6.90 (2H, d., *J* 8 Hz, aromatic), 7.01 (1H, d., *J* 7.5 Hz, amide NH), 7.06 (1H, s., 1 of 1° amide NH), 7.35 (1H, s., 1 of 1° amide NH), 7.57 (1H, d., *J* 7.2 Hz, urethane NH) and 9.07 (1H, s., aromatic OH); δ_C (75.74 MHz, d_6-DMSO) 20.51 & 22.72 ($(CH_3)_2$), 24.12 ($\underline{C}H(CH_3)_2$), 28.09 ($(CH_3)_3C$), 36.83 (CH_2 Leu), 40.65 (CH_2 Ar), 53.6 (Tyr CH), 56.1 (Leu CH), 78.14 ($(CH_3)_3\underline{C}$), 114.74, 127.55 & 129.99 (aromatic), 152.71 (urethane CO) and 171.83 & 172.68 (CO); *m/z* (CI) 394 ([*M*+H]$^+$, 14%), 338 ([(*M*-$(CH_3)_3CO$) + NH_4]$^+$, 58), 294 ([(*M* -Boc) + NH_4]$^+$, 100) and 107 ($CH_2C_6H_4OH^+$, 12).

(2S)-Leucyl-(2S)-tyrosyl amide hydrochloride (11).-

Peptide (**10**) (100 mg, 0.254 mmol) was dissolved in ethyl acetate (10 ml), cooled to 0 °C and dry hydrogen chloride gas bubbled into the solution for 10 min. The solution was stirred at 0 °C for 1 h. and then allowed to warm to room temperature over 15 min. The solvent was removed under reduced pressure to give a white solid that was washed with diethyl ether (2 x 5 ml) and dried under vacuum to give the product as a white solid (65.3 mg, 79%) of m.p. 110-112 °C. *m/z* (Found: $(M + H)^+$ 294.1810. $C_{15}H_{23}N_3O_3$ requires 294.1817); $[\alpha]_D^{20}$ -23.1° (c 2 in MeOH); ν_{max} (Nujol) 3541 (OH), 3470-3210 (NH), 1683 & 1676 (amide CO) and 758 cm^{-1} (aromatic); δ_H (270 MHz, d_6-DMSO) 0.83 (3H, d., *J* 6.0 Hz, 1 of $(C\underline{H}_3)_2CH$), 0.87 (3H, d., *J* 5.6 Hz, 1 of $(C\underline{H}_3)_2CH$), 1.53 (2H, m., $CHC\underline{H}_2$), 1.61 (1H, m, $C\underline{H}CH_2$), 2.60 (1H, d.d., J_{AB} 14.1 Hz, J_{AX} 8.5 Hz, 1 of CH_2Ph), 2.88 (1H, d.d., J_{AB} 14.1 Hz, J_{BX} 5.4 Hz, 1 of CH_2Ph), 3.73 (1H, t., *J* 6.8 Hz, NHCH Leu), 4.29 (1H, m., NHCH Tyr), 6.60 (2H, d., *J* 7.5 Hz, aromatic), 7.03 (2H, d., *J* 7.5 Hz, aromatic), 7.04 (1H, s., 1 of 1° amide NH_2), 7.55 (1H, s., 1 of 1° amide NH_2), 8.09 (2H, s., amine NH_2), 8.68 (1H, d., *J* 8.14 Hz, 2° amide NH) and 9.07 (1H, s., aromatic OH); δ_C (50.31 MHz, ^2H_2O) 21.66 & 22.35 ($(\underline{C}H_3)_2$), 24.48 ($\underline{C}H(CH_3)_2$), 36.82 (CH_2 Leu), 41.01 (CH_2 Tyr), 53.62 (Tyr CH), 56.14 (Leu CH), 116.66, 128.90 & 131.55 (aromatic), 155.23 (urethane CO) and 170.81 & 175.94 (amide CO); *m/z* (FAB) 293 ($[M$-HCl$]^+$, 1.8%), 208 ($[M$+H-$NH_2CHCH_2CH(CH_3)_2]^+$, 2.8), 187 ($[M$+H-$CH_2C_6H_4OH]^+$, 5.2), 136 ($NH_3CHCHC_6H_4OH^+$, 6.3) and 86 ($NH_3CHCH_2(CH_3)_2^+$, 100).

N-Acetyl-(2S)-phenylalanine (12).- Method 1

(2S)-Phenylalanine (2.25 g, 13.8 mmol) was dissolved in water (30 ml) and the pH adjusted to 11.5 with 2 M sodium hydroxide. Acetic anhydride (3.6 ml) was dissolved in benzene (30 ml), the aqueous phase added and the two phase mixture shaken vigorously for 15 min., then stirred at room temperature for 2 h. Ethyl acetate (15 ml) was then added and the two phases separated, the pH of the aqueous phase adjusted to 2 with 2 M HCl and extracted with ethyl acetate (3 x 15 ml). The organic extracts were combined, dried ($MgSO_4$) and the solvent removed under reduced pressure to give a white solid, (1.77 g, 62%). This was identical in all respects to the

sample produced by method 2.

Method 2[288]

N-Methoxydiacetamide, the acetylating agent, was prepared as follows; Triethylamine (12 g, 120 mmol) was mixed with O-methyl hydroxylamine hydrochloride (2.5 g, 30 mmol) in dichloromethane (25 ml) with cooling in ice. Acetic anhydride (7.6 g, 75 mmol) was then added dropwise during 30 min. and the reaction mixture stirred overnight. The solution was then filtered and the filtrate washed with 5% sodium bicarbonate solution (3 x 15 ml), dried ($MgSO_4$) and the solvent removed under reduced pressure to give an oil. Distillation under reduced pressure gave the product as a colourless oil (72-74 °C, 12 mm Hg) (2.91 g, 74%). ν_{max} (Nujol) 1720 cm^{-1} (amide CO); δ_H (90 MHz, C^2HCl_3) 2.31 (6H, s., CH_3CO) and 3.74 (3H, s., OCH_3); *m/z* (EI) 131 (M^+, 1.9%), 89 ([M+H-CH_3CO]$^+$, 38.2) and 43 (CH_3CO^+, 100).

To a stirred solution of (2S)-phenylalanine (169 mg, 1 mmol) in dioxane (4 ml) was added 2 M sodium hydroxide (0.5 ml), followed by N-methoxydiacetamide (262 mg, 2 mmol) at room temperature. After 27 h. the reaction was diluted with water (4 ml), the pH adjusted to 4 with glacial acetic acid and the aqueous phase extracted with ethyl acetate (3 x 20 ml). The organic extracts were combined, dried ($MgSO_4$) and the solvent removed under reduced pressure to give a white solid. The solid was suspended in diethyl ether, filtered, washed with diethyl ether (15 ml) and dried to give the product as a white solid (158 mg, 71%) of m.p. 170-172 °C (lit.[436] 171-3 °C). *m/z* (Found: (M + H)$^+$ 208.0974. $C_{11}H_{14}NO_3$ requires 208.0974); $[\alpha]_D^{20}$ +48° (c 1 in MeOH) (lit.[436] $[\alpha]_D^{20}$ +46.5° (c 2 in EtOH)); ν_{max} (Nujol) 3329 (NH), 1688 (acid CO), 1617 (amide CO), 1552 (amide II) and 752 cm^{-1} (aromatic); δ_H (90 MHz, d_6-DMSO) 1.82 (3H, s., CH_3CO), 3.02 (1H, d.d., J_{AX} 7.9 Hz, J_{AB} 14.05 Hz, 1 of CH_2Ph), 3.05 (1H, d.d., J_{BX} 6.7 Hz, J_{AB} 14.05 Hz, 1 of CH_2Ph), 4.42 (1H, m., CH), 7.29 (5H, s., aromatic) and 8.20 (1H, d., J 8.2 Hz, NH); δ_C (50.31 MHz, C^2HCl_3) 23.59 (CH_3), 38.29 (CH_2), 53.64 (CH), 127.63-129.72 (aromatic), 136.35 (quat. aromatic) and 170.32 & 172.68 (CO); *m/z* (EI) 207 (M^+, 1.62%), 163 ([M+H-CO_2H]$^+$, 4.3), 148 ([M-CH_3CONH_2]$^+$, 100), 91 ($PhCH_2^+$, 71.55) and 74 ($HO_2CCHNH_2^+$, 38.0).

Methyl (2S)-phenylalaninate hydrochloride (13).-

To a stirred solution of (2S)-phenylalanine (4.20 g, 25 mmol) in dry methanol (20 ml) at 0 °C was added thionyl chloride (4.02 ml, 55 mmol) dropwise with stirring. The solution was then refluxed for 2 h. and the solvent removed under reduced pressure to give a cream solid. Recrystallisation from methanol/ diethyl ether gave the product as white crystals (4.09 g, 76.5%) of m.p. 159-162 °C (lit.[437] 158-162 °C). *m/z* (Found: $(M + H)^+$ 180.1024. $C_{10}H_{14}NO_2$ requires 180.1024); $[\alpha]_D^{20}$ -6.4° (c 1 in H_2O) (lit.[437] $[\alpha]_D^{20}$ -4.6 (c 5 in H_2O)); ν_{max} (Nujol) 3153 (NH), 1745 (ester CO), 1583 (NH bend) and 763 cm^{-1} (aromatic); δ_H (90 MHz, ^2H_2O) 3.04 (1H, d.d., J_{AB} 14.1 Hz and J_{AX} 5.5 Hz, CH_2Ph), 3.10 (1H, d.d., J_{AB} 14.1 Hz and J_{BX} 7.6 Hz), 3.69 (3H, s., OMe), 4.38 (1H, d.d., J_{BX} 7.6 Hz, J_{AX} 5.5 Hz, NHCH) and 7.28 (5H, m., Ph); *m/z* (EI) 179 ([*M*-HCl]$^+$, 1.4%), 120 ([*M*-CO_2Me]$^+$, 89.3), 91 ($PhCH_2^+$, 31.7) and 87 ([*M*-$PhCH_2$-HCl]$^+$, 100).

Methyl N-acetyl-(2S)-phenylalaninate (14).- Method 1[438]

N-Acetyl (2S)-phenylalanine (**12**) (207 mg, 1 mmol) was suspended in diethyl ether (10 ml) and cooled to 0 °C. Excess ethereal diazomethane (0.2 g, 4.8 mmol) was added with stirring and the mixture stirred at room temperature for 1 h. Excess diazomethane was then removed in a stream of nitrogen and the solvent removed under reduced pressure to give a yellowish solid, which was recrystallised from diethyl ether/ light petroleum to give the product as white crystals (175 mg, 79%). In all respects it was identical with the sample from method 3.

Method 2

The ester (**13**) (431 mg, 2 mmol) was dissolved in glacial acetic acid (10 ml) and distilled acetic anhydride (224 mg, 2.2 mmol) added. The mixture was stirred at room temperature for 96 h., the solvent was removed under reduced pressure to give a white solid that was recrystallised from ethyl acetate/ light petroleum to give the product as a white solid (350 mg, 68%). This was identical in all respects with the sample from method 3.

Method 3

Ester (**13**) (1.35 g, 6.25 mmol) was suspended in chloroform (35 ml), triethylamine (871 µl, 6.25 mmol) added with stirring and the solution cooled to -40 °C. Acetyl chloride (440 µl, 6.25 mmol) was added dropwise over 20 min., and the solution stirred for 40 min. at room temperature. The solvent was removed under reduced pressure and the residue triturated with ethyl acetate (50 ml). The resulting solution was washed with 0.5 M HCl (20 ml), 5% sodium bicarbonate solution (20 ml), water (20 ml) and brine (20 ml). The solution was dried ($MgSO_4$) and the solvent removed under reduced pressure to give a white solid that was recrystallised from ethyl acetate/ light petroleum to give the product (1.14 g, 83%) as white crystals of m.p. 90-91 °C (lit.[438] 90-91 °C). (Found: C, 64.89; H, 6.68; N, 6.28. Calc. for $C_{12}H_{15}NO_3$: C, 65.15; H, 6.79; N, 6.28); $[\alpha]_D^{20}$ +14.2° (c 2 in MeOH) (lit.[438] $[\alpha]_D^{20}$ +16.5° (c 2 in MeOH)); ν_{max} (Nujol) 3328 (NH), 1699 (ester CO), 1621 (amide CO), 1533 (amide II), 1243 (C-O) and 746 cm^{-1} (aromatic); δ_H (360 MHz, C^2HCl_3) 1.99 (3H, s., CH_3CO), 3.06 (1H, d.d., J_{AB} 13.9 Hz and J_{AX} 5.8 Hz, 1 of CH_2Ph), 3.12 (1H, d.d., J_{AB} 13.9 Hz and J_{BX} 8.1 Hz, 1 of CH_2Ph), 3.73 (3H, s., OCH_3), 4.88 (1H, d.d.d., J_{BX} 8.1 Hz, J_{NH} 7.2 Hz, J_{AX} 5.8 Hz , NHC<u>H</u>), 6.02 (1H, br. d., J 7.2 Hz, NH) and 7.05-7.35 (5H, m., Ph); δ_C (90.56 MHz, C^2HCl_3) 23.2 (<u>C</u>H_3CO), 37.9 (CH_2Ph), 52.3 (OCH_3), 53.2 (CHNH), 127.2, 128.7 & 129.3 (aromatic), 135.9 (quat. aromatic) and 169.7 & 172.2 (CO); *m/z* (EI) 221 (M^+, 3.1%), 178 ([*M*-CH_3CO]$^+$, 2.2), 162 ([*M*-CO_2Me]$^+$, 100), 131 ([*M*+H-CH_2Ph]$^+$, 33.6), 120 ($NH_2CHCH_2Ph^+$, 45.0), 88 ($NH_2CHCO_2Me^+$, 64.5) and 43 (CH_3CO^+, 43.8).

N-*tert*-Butoxycarbonyl-(2S)-leucinol (15).- Method 1

N-*tert*-Butoxycarbonyl-(2S)-leucine hydrate (**9**) (400 mg, 1.75 mmol) was dried as for compound (**10**) and dissolved in dry THF (10 ml). The solution was cooled to -20 °C and triethylamine (156 mg, 1.75 mmol) and 2-methylpropylchloroformate (228 µl, 1.75 mmol) were added consecutively. The solution was stirred for 15 min., then filtered directly into a 0 °C solution of sodium borohydride (100 mg, 2.5 mmol) in water (10 ml). The solution was allowed to warm to room temperature and stirred for 3 h. The mixture was acidified to pH 3 with 2 M HCl and extracted with diethyl ether (4 x 10 ml). The ethereal extracts were combined, dried ($MgSO_4$) and evaporated

under reduced pressure. The resulting yellow oil was purified by flash chromatography (diethyl ether/ light petroleum, 7:3) to give the product as a clear oil (273 mg, 72%). ν_{max} (thin film) 3349 (OH), 2956 (NH), 2790-2650 (CH), 1685 (urethane CO) and 1173 cm^{-1} (amide II); δ_H (270 MHz, C^2HCl_3) 0.91 (6H, d., *J* 7 Hz, $(C\underline{H}_3)_2CH$), 1.34 (2H, m., $C\underline{H}_2CH$), 1.47 (9H, s., $(CH_3)_3C$), 1.49 (1H, m., $C\underline{H}CH_2$), 3.54 (1H, m., $C\underline{H}_2OH$), 4.30 (1H, m., NHC$\underline{H}$), 4.62 (1H, s., OH) and 4.94 (1H, d., *J* 7.7 Hz, NH); *m/z* (EI) 217 (M^+, 1.5%), 186 ([*M*-CH_2OH]$^+$, 3.9), 130 ([*M*+H-CH_2OH-$(CH_3)_3C$]$^+$, 33.8), 86 ([*M*+H-CH_2OH-Boc]$^+$, 21.4) and 57 ($(CH_3)_3C^+$, 100).

Method 2[439]

N-*tert*-Butoxycarbonyl-(2S)-leucine hydrate (**9**) (205 mg, 0.887 mmol) was dried as for compound (**10**), dissolved in dry THF (4 ml) and cooled to -10 °C. N-methyl morpholine (89.6 mg, 0.887 mmol) and ethyl chloroformate (95.8 mg, 0.887 mmol) were added and the solution stirred for 10 min. Sodium borohydride (100 mg, 2.66 mmol) was then added in one portion and dry methanol (8 ml) added dropwise over the next 10 min. at 0 °C. After 10 min. the reaction was neutralised with 1 M HCl. The organic solvents were removed under reduced pressure and the resulting aqueous layer extracted with ethyl acetate (3 x 8 ml). The organic layer was washed successively with 1 M HCl (3 ml), water (2 x 8 ml), 5% sodium bicarbonate (4 ml) and water (2 x 8 ml) and then dried ($MgSO_4$). The solvent was evaporated under reduced pressure to give a clear oil, which was further purified by chromatography on flash silica (diethyl ether/ light petroleum, 7:3) to give a clear oil (133 mg, 69%) identical to the sample prepared by Method 1.

(2S)-Leucinol hydrochloride (16).-

This was synthesised in an identical manner to the hydrochloride (**11**), starting with (**15**) (217 mg, 1 mmol), to give the product as a white solid (125 mg, 82%) of m.p. 156-8 °C. $[\alpha]_D^{20}$ +0.9° (c 3 in MeOH); ν_{max} (Nujol) 3375 (OH), 3100 (NH_3^+) and 1375 cm^{-1} (OH bend); δ_H (270 MHz, ^2H_2O) 0.71 (3H, d., *J* 6.2 Hz, 1 of $(C\underline{H}_3)_2CH$), 0.73 (3H, d., *J* 6.2 Hz, 1 of $(C\underline{H}_3)_2CH$), 1.27 (1H, m., $C\underline{H}(CH_3)_2$), 1.54 (2H, m., $C\underline{H}_2CH$), 3.37 & 3.58 (2H, m., $C\underline{H}_2OH$) and 3.79 (1H, t., *J* 7.8 Hz, CH); *m/z* (EI) 157 ([*M*+H]$^+$, 5.8%), 86 ([*M*-HCl-CH_2OH]$^+$, 100), 60 ($NH_2CHCH_2OH^+$, 19.6) and 44 ($CHCH_2OH^+$,

66).

Methyl (2S)-valinate hydrochloride (17).-

This was made in an identical manner to (**13**), using (2S)-valine (1.17 g, 10 mmol). Recrystallisation from methanol/ diethyl ether gave the product as white crystals (1.09 g, 69%) of m.p. 164-166 °C (lit.[440] 166-168 °C). *m/z* (Found: $(M + H)^+$ 207.0896. $C_{11}H_{13}NO_3$ requires 207.0895); $[\alpha]_D^{20}$ +15.0° (c 1 in H_2O) (lit.[440] $[\alpha]_D^{20}$ +15.5° (c 1 in H_2O)); ν_{max} (Nujol) 3324 (NH), 1738 (ester CO), 1506 (NH bend) and 1243 cm^{-1} (C-O); δ_H (90 MHz, ^2H_2O) 0.91 (6H, d., *J* 7.5 Hz, $(C\underline{H}_3)_2CH$), 2.25 (1H, m., $(CH_3)_2C\underline{H}$), 3.69 (3H, s., OCH_3) and 3.89 (1H, d., *J* 7.2 Hz, C<u>H</u>NH); *m/z* (EI) 131 (M^+, 1.2%), 88 ($[M\text{-}CH(CH_3)_2]^+$, 12.3), 72 ($[M\text{-}CO_2Me]^+$, 100) and 36 ($NH_2CHCH_2^+$, 34.6).

Methyl N-acetyl-(2S)-phenylalanyl-(2S)-valinate (18).-

N-Acetyl-(2S)-phenylalanine (**12**) (208 mg, 1 mmol) was dissolved in dry DMF (5 ml) at -5°C and N-methyl morpholine (110 µl, 1 mmol) and 2-methylpropylchloroformate (130 µl, 1 mmol) added with stirring. The mixture was maintained at -5 °C for 20 min. and the ester (**17**), (131 mg, 1 mmol) added in DMF (5 ml) containing N-methyl morpholine (110 µl, 1 mmol). The mixture was then allowed to warm to room temperature and stirred overnight. Water (10 ml) was added and the mixture extracted with ethyl acetate (3 x 10 ml), which was then washed with 0.5 M HCl (3 x 5 ml), 5% sodium bicarbonate solution (3 x 5 ml) and saturated brine (10 ml). The organic extract was dried ($MgSO_4$) and the solvent removed under reduced pressure to give a clear oil, which was crystallised from methanol/ light petroleum to give the product as white crystals (158 mg, 49%) of m.p. 127-130 °C. *m/z* (Found: $(M + H)^+$ 321.1814. $C_{17}H_{25}N_2O_4$ requires 321.1814); $[\alpha]_D^{20}$ -9.2° (c 2 in $CHCl_3$); ν_{max} (Nujol) 3019 (NH), 1738 (ester CO), 1695 (acetyl amide CO), 1656 (2° amide), 1516 (NH bend) and 1223 cm^{-1} (C-O); δ_H (360 MHz, C^2HCl_3) 0.81 (6H, d., *J* 8.4 Hz, $(C\underline{H}_3)_2$), 1.94 (3H, s., CH_3CO), 2.12 (1H, m., $C\underline{H}(CH_3)_2$), 2.78 (1H, d.d., J_{AB}

12.8 Hz, J_{AX} 7.6 Hz, 1 of CH_2Ph), 2.97 (1H, d.d., J_{AB} 12.8 Hz, J_{BX} 5.3 Hz, 1 of CH_2Ph), 3.69 (3H, s., OCH_3), 4.40 (1H, m., Phe NHC<u>H</u>), 4.92 (1H, m., Val NHC<u>H</u>), 6.94 (1H, d., *J* 8.2 Hz, Val NH), 7.12 (1H, d., *J* 8.4 Hz, Phe NH) and 7.31 (5H, m., aromatic); δ_C (50.31 MHz, C^2HCl_3) 15.3 & 16.7 ($(CH_3)_2$), 20.7 (<u>C</u>H_3CO), 28.8 (<u>C</u>$H(CH_3)_2$), 36.4 (CH_2Ph), 49.7 (Val NHCH), 52.2 (Phe NHCH), 55.3 (CH_3O), 124.8, 126.7 & 127.2 (aromatic) and 172.8, 174.5 & 175.2 (CO); *m/z* (EI) 320 (M^+, 18.9%), 261 ([*M*-CO_2Me]$^+$, 34.3), 187 ([*M*-CH_3CO-CH_2Ph]$^+$, 34.4), 120 ($NH_2CHCH_2Ph^+$, 100) and 91 ($PhCH_2^+$, 23.6).

N-Acetyl-(2S)-phenylalanyl-(2S)-valine (19).-

Methyl ester (**18**) (200 mg, 0.625 mmol) was dissolved in a 5% solution of water in methanol (10 ml), containing potassium carbonate (95 mg, 0.68 mmol). The mixture was stirred at room temperature for 48 h. The solution was then washed with chloroform (5 ml), acidified to pH 2 with 0.5 M HCl and extracted with chloroform (3 x 5 ml). The organic extract was washed with water (8 ml), dried ($MgSO_4$) and the solvent removed under reduced pressure to give a cream solid, which was recrystallised from methanol/ diethyl ether to give the product as white crystals (155 mg, 81%) of m.p. 103-107 °C. ν_{max} (Nujol) 3419 (NH), 3013-2850 (COOH),1708 (acid CO), 1695 (acetyl amide CO), 1656 (2° amide), 1516 (NH bend) and 754 cm^{-1} (aromatic); δ_H (90 MHz, C^2HCl_3) 0.85 (6H, m., ($C\underline{H}_3)_2CH$), 1.88 (3H, s., CH_3CO), 2.01 (1H, m., $C\underline{H}(CH_3)_2$), 2.68 (1H, d.d., J_{AB} 13.2 Hz, J_{AX} 7.4 Hz, 1 of CH_2Ph), 2.93 (1H, d.d., J_{AB} 13.2 Hz, J_{BX} 5.2 Hz, 1 of CH_2Ph), 4.42 (1H, m., Phe NHC<u>H</u>), 5.01 (1H, m., Val NHC<u>H</u>), 7.31 (5H, m., aromatic), 7.43 (1H, m., Val NH) and 8.84 (1H, br. s., CO_2H); δ_C (50.31 MHz, C^2HCl_3) 18.38 & 19.32 (Val CH_3), 23.21 (<u>C</u>H_3CO), 31.32 (<u>C</u>$H(CH_3)_2$), 55.04 & 58.16 (NHCH x2), 127.42, 129.01 & 129.88 (aromatic) and 171.50, 172.35 & 173.93 (CO).

Methyl (2S)-alaninate hydrochloride (20).-

This was prepared exactly as for (**13**), using (2S)-alanine (2.25 g, 25 mmol). Recrystallisation from dry ethanol/ dry diethyl ether gave the product as hygroscopic

white crystals (3.12 g, 89%) of m.p. 108-110 °C (lit.[441] 109-111 °C). *m/z* (Found: (*M* + H)$^+$ 104.0711. $C_4H_{10}NO_2$ requires 104.0711); $[\alpha]_D^{20}$ +6.1° (c 2 in MeOH); ν_{max} (Nujol) 3418 (NH) and 1742 cm^{-1} (ester CO); δ_H (90 MHz, ^2H_2O) 1.43 (2H, d., *J* 7.5 Hz, CHC<u>H</u>$_3$), 3.72 (3H, s., OCH_3) and 4.08 (1H, q., *J* 7.5 Hz, CH); δ_C (50.31 MHz, ^2H_2O) 17.87 (<u>C</u>H$_3$CH), 51.84 (CH), 56.95 (OCH_3) and 174.89 (CO); *m/z* (EI) 103 (*M*$^+$, 1.1%), 88 ([*M*-CH_3]$^+$, 3.02), 44 ($H_2NCHCH_3^+$, 100) and 36 (HCl, 23.77).

Methyl (2S)-leucinate hydrochloride (21).-

This was prepared in an identical manner to the (2S)-phenyalanine analogue (**13**), using (2S)-leucine (655 mg, 5 mmol). The product was obtained as white crystals, after recrystallisation from methanol/ diethyl ether, of m.p. 150-153 °C (lit.[442] 148-150 °C). *m/z* (Found: (*M* + H)$^+$ 146.1181. $C_7H_{16}NO_2$ requires 146.1181); $[\alpha]_D^{20}$ -15.2° (c 2 in MeOH) (lit.[442] $[\alpha]_D^{20}$ -13.4° (c 5 in H_2O)); ν_{max} (Nujol) 3484 (NH), 1737 (ester CO) and 1591 cm^{-1} (NH_3^+); δ_H (90 MHz, ^2H_2O) 0.82 (6H, d., *J* 5.7 Hz, (C<u>H</u>$_3$)$_2$), 1.3-1.8 (1H, m., CH_2CH), 3.68 (3H, s., OMe) and 4.01 (1H, t., *J* 6 Hz, H_3NC<u>H</u>); δ_C (50.31 MHz, ^2H_2O) 23.72 & 24.28 ((<u>C</u>H$_3$)$_2$CH), 26.72 ((CH$_3$)$_2$<u>C</u>H), 41.58 (<u>C</u>H$_2$CH), 54.20 ($CHNH_3$), 56.27 (CH_3O) and 174.16 (ester CO); *m/z* (EI) 146 (*M*$^+$, 17.8%) and 86 ([*M*-CO_2Me]$^+$, 100).

(2S)-3-Methyloxiranemethanol (22).-[309]

Crushed, activated 3 Å molecular sieves (3.0 g) were added to a flame-dried flask under nitrogen, followed by dry dichloromethane (200 ml). The flask was then cooled to -20 °C with stirring and (2S)-(+)-diisopropyl tartrate (1.42 g, 6 mmol), *E*-2-buten-1-ol (7.21 g, 100 mmol, stored over 3Å molecular sieves) and titanium isopropoxide (1.42 g, 5.0 mmol) were added sequentially. Stirring was maintained for 15 min. at -20 °C, then a 3.0 M solution of *tert*-butyl hydroperoxide (67 ml, 200 mmol) in dry dichloromethane (200 ml) was added and the mixture stirred at -20 °C for 2 h. The reaction was quenched by the slow addition of tributylphosphine (20.2 g, 100 mmol) at -20 °C. The mixture was checked by t.l.c. (40 % ethyl acetate: light petroleum) to ensure all the hydroperoxide had been reduced. Citric acid monohydrate (1.05 g) in 10% acetone in diethyl ether (150 ml) was added and the mixture stirred at room

temperature for 30 min. After filtration through a pad of Celite the mixture was concentrated under reduced pressure to yield a yellow viscous oil, which was distilled under reduced pressure (18 mm Hg, 81-83 °C) to yield a clear colourless oil, which proved to be very difficult to purify any further. δ_H (300 MHz, C^2HCl_3) 1.36 (3H, d., *J* 5.6 Hz, $CH_3C\underline{H}$), 2.91 (1H, d.d.d., *J* 2.9, 3.0, 3.9 Hz, $C\underline{H}CH_2OH$), 3.04 (1H, d.q., *J* 2.9, 5.6 Hz, $CH_3C\underline{H}$), 3.61 (1H, d.d., *J* 3.7, 12.9 Hz, 1 of $C\underline{H}_2OH$) and 3.89 (1H, d.d., *J* 3.0, 12.9 Hz, 1 of $C\underline{H}_2OH$).

N-9-Fluorenylmethyloxycarbonyl-(2S)-threonine (23).-[443]

To a stirred suspension of (2S)-threonine (2.26 g, 19 mmol) in 10% sodium carbonate solution (30 ml) and dioxan (20 ml) at 0°C was added dropwise 9-fluorenylmethylchloroformate (4.9 g, 19 mmol) in dioxan (20 ml). After 1 h. the solution was warmed to room temperature and stirred for 4 h. The mixture was then poured into ice-water (250 ml) and extracted with diethyl ether (3 x 120 ml). The aqueous layer was cooled in an ice-bath during acidification to pH 2, causing a yellow oil to separate that was taken up in ethyl acetate (250 ml). The organic layer was washed with 0.1 M HCl (50 ml), water (50 ml) and brine (50 ml). The organic layer was dried ($MgSO_4$) and the solvent removed under reduced pressure to give an off-white solid. Trituration with light petroleum gave the product as a cream solid (4.79 g, 74%) of m.p. 102-106 °C that was refractory to recrystallisation. *m/z* (Found: $(M + H)^+$ 342.1341. $C_{19}H_{20}NO_5$ requires 342.1341); $[\alpha]_D^{20}$ -6.4° (c 2 in $CHCl_3$); ν_{max} ($CHCl_3$) 3587-3340 (OH), 3432 (NH), 3020-2930 (aromatic), 1742 (carboxyl CO), 1723 (urethane CO) and 1520 cm^{-1} (amide II); δ_H (90 MHz, C^2HCl_3) 1.21 (3H, d., *J* 7.1 Hz, CH_3), 4.18 (1H, t., *J* 7.5 Hz, Fmoc CH), 4.21 (4H, m., Fmoc CH_2, C-3 and C-2 H), 5.94 (1H, d., *J* 8.4 Hz, NH) and 7.24-7.7 (aromatic); δ_C (50.31 MHz, C^2HCl_3) 20.41 ($\underline{C}H_3CH$), 47.61 (CH_2), 59.92 (CHOH), 67.73 (Fmoc CH), 68.2 (CHNH), 120.55, 125.69, 127.65 & 128.29 (aromatic), 141.8 & 144.21 (quat. aromatic), 157.54 (urethane CO) and 172.45 (carboxyl CO); *m/z* (CI) 359 ($[M+NH_4]^+$, 89%), 342 ($[M+H]^+$, 96), 178 ([9-fluorene]$^+$, 100) and 120 ([threonine+H]$^+$, 47).

Methyl 9-Fluorenylmethyloxycarbonyl-(2S)-threoninate (24).-

To a stirred solution of Fmoc (2S)-threonine (**23**), (1.62 g, 4.75 mmol) in diethyl ether (50 ml) at 0 °C was added excess ethereal diazomethane (0.5 g, 12 mmol) and the solution stirred at room temperature for 1 h. Unreacted diazomethane was removed in a stream of nitrogen and the solvent removed under reduced pressure to give a yellowish solid. Low pressure column chromatography on t.l.c. silica (ethyl acetate) gave a white solid. Recrystallisation from diethyl ether/ light petroleum gave the product as a white solid (1.21 g, 72%) of m.p. 118-119 °C. *m/z* (Found: $(M + H)^+$ 356.1498. $C_{11}H_{13}NO_3$ requires 356.1498); $[\alpha]_D^{20}$ -12.8° (c 1 in $CHCl_3$); ν_{max} ($CHCl_3$) 3495 (OH), 1758 (ester CO), 1723 (urethane CO), 1493 (aromatic) and 1212 cm^{-1} (aromatic); δ_H (300 MHz, C^2HCl_3) 1.22 (3H, d., *J* 7.1 Hz, CH_3), 3.72 (3H, s., OCH_3), 4.21 (1H, t., *J* 7.4 Hz, Fmoc CH), 4.3-4.5 (4H, m., Fmoc CH_2, C-2 and C-3 H), 5.90 (1H, d., *J* 9 Hz, -NH) and 7.1-7.7 (8H, m., aromatic); δ_C (75.47 MHz, C^2HCl_3) 20.15 (CH_3), 47.53 (CH_2), 52.91 (OCH_3), 59.48 (CHOH), 67.52 (Fmoc CH), 68.21 (CHNH), 120.41, 125.12, 127.24 & 127.66 (aromatic), 142.77 & 144.12 (quat. aromatic), 156.80 (urethane CO) and 174.98 (ester CO); *m/z* (CI) 356 ($[M + H]^+$, 6.8%), 179 ([9-fluorene+ H]$^+$, 97) and 134 ($[M$-Fmoc+H$]^+$, 100).

Methyl (2R,3S)-(9-Fluorenylmethyloxycarbonyl)-2-amino-3-bromobutyrate (25).-[306]

Carbon tetrabromide (3.36 g, 10.1 mmol) and triphenylphosphine (2.66 g, 10.1 mmol) were dissolved in dry benzene (60 ml) at room temperature and the methyl ester (**24**) (2.5 g, 7 mmol) added in portions over 30 min. The mixture was stirred for a further 90 min., filtered and the pad washed with dry benzene (30 ml). The benzene was removed under reduced pressure to give a yellow oil which was purified by flash column chromatography (ethyl acetate/ chloroform (3:1)), to give an oil. Further flash column chromatography (chloroform) gave a clear oil. Crystallisation from diethyl ether/ light petroleum gave the product as a white solid (2.6 g, 61%) of m.p. 147-150 °C. $[\alpha]_D^{20}$ +7.3 ° (c 2 in $CHCl_3$); ν_{max} (Nujol) 3415 (NH), 1745 (ester CO), 1718 (urethane CO) and 1507 cm^{-1} (aromatic); δ_H (300 MHz, C^2HCl_3) 1.77 (3H, d., *J* 7.0 Hz, CH_3), 3.72 (3H, s., OCH_3), 4.21 (1H, t., *J* 7.2 Hz, Fmoc CH), 4.37 (3H, m., Fmoc

CH_2 and C-2 H), 4.59 (1H, m., C-3 H), 5.72 (1H, d., *J* 8.9 Hz, NH) and 7.1-7.7 (8H, m., aromatic); δ_C (75.47 MHz, C^2HCl_3) 23.68 (CH_3), 48.02 (CH_2), 50.12 (OCH_3), 53.61 (C-2), 60.65 (3-CH), 68.22 (Fmoc CH), 120.94, 125.96, 127.95 and 128.60 (aromatic), 142.23 & 144.53 (quat. aromatic), 156.41 (urethane CO) and 169.98 (ester CO).

Methyl N-*tert*-butoxycarbonyl-(2S)-leucyl-(2S)-leucinate (26).-[444]

To a stirred solution of N-*tert*-butoxycarbonyl-(2S)-leucine hydrate (**9**) (2.31 g, 10 mmol), dried as for (**10**), in dry THF (25 ml) was added N-methylmorpholine (1.10 ml, 10 mmol) and the solution cooled to -15 °C. 2-methylpropylchloroformate (1.36 ml, 10 mmol) was added with stirring and the solution stirred for 1 min. Methyl ester (**21**) (1.81 g, 10 mmol) in DMF (15 ml) containing N-methylmorpholine (1.12 ml, 10 mmol) was then added in one portion. After 1 min. at -15 °C and 5 min. at room temperature the hydrochlorides were filtered off and the THF removed under reduced pressure. Addition of water to the residual liquid precipitated a white solid, which was recrystallised from methanol/ water to give the product as a white solid (2.83 g, 79%) of m.p. 134-136 °C (lit.[445] 132-133 °C). *m/z* (Found: $(M + H)^+$ 359.2547. $C_{18}H_{35}N_2O_5$ requires 359.2546); $[\alpha]_D^{20}$ -48.9° (c 1 in MeOH) (lit.[445] $[\alpha]_D^{20}$ -50.4° (c 1 in MeOH)); ν_{max} ($CHCl_3$) 3020-2960 (CH_2, CH), 1741 (ester CO), 1705 (urethane CO), 1679 (amide CO) and 1215 cm^{-1} (C-O); δ_H (360 MHz, C^2HCl_3) 0.97 (12H, m., $(C\underline{H}_3)_2$ x 2), 1.44 (9H, s., $(CH_3)_3C$), 1.46-1.80 (6H, m., $C\underline{H}_2C\underline{H}$ x 2), 4.10 (1H, m., NHC<u>H</u> Leu2), 4.62 (1H, m., NHC<u>H</u> Leu1), 4.89 (1H, d., *J* 7.3 Hz, NH Leu2) and 6.47 (1H, d., *J* 8.1 Hz, NH Leu1); δ_C (50.31 MHz, C^2HCl_3) 22.48 & 23.42 ($(CH_3)_2$ x2), 25.31 (CH), 28.85 ($(\underline{C}H_3)_3C$), 41.52 and 42.14 (CH_2 x2), 51.26 (NHCH Leu 1), 52.7 (OCH_3), 54.55 (NHCH Leu 2), 80.13 ($(CH_3)_3\underline{C}$) and 172.81 & 173.6 (CO x2); *m/z* (EI) 358 (M^+, 0.33%), 302 ($[M+H-(CH_3)_3C]^+$, 4), 299 ($[M-CO_2Me]^+$, 0.5), 130 ($[CH(CO_2Me)CH_2(CH_3)_2+H]^+$, 79.1), 86 ($NH_2CHCH_2CH(CH_3)_2^+$, 100) and 57 ($(CH_3)_3C^+$, 66.3).

N-*tert*-Butoxycarbonyl-(2S)-leucyl-(2S)-leucine (27).-[446]

The methyl ester (**26**) (3.58 g, 10 mmol) was dissolved in methanol (20 ml), the flask surrounded by a water bath at room temperature and 1 M sodium hydroxide (12 ml) added with stirring. The resulting precipitate was dissolved in the minimum quantity of methanol and the solution stirred at room temperature for 2 h. The methanol was removed under reduced pressure and the aqueous solution stirred in an ice-bath during acidification to pH 2. The solution was kept at 0 °C for a further 2 min. and the resulting white solid collected by filtration. Recrystallisation from methanol/ water gave the product, after drying over phosphorous pentoxide, as a white solid (3.09 g, 90%) of m.p. 137-138 °C. *m/z* (Found: $(M + H)^+$ 345.2387. $C_{17}H_{33}N_2O_5$ requires 345.2389); $[\alpha]_D^{20}$ -18.2° (c 2 in $CHCl_3$); ν_{max} (Nujol) 3431 (NH), 3336 (NH), 1705 (acid CO), 1684 (urethane CO), 1637 (amide CO), 1536 (amide II) and 1275 cm^{-1} (C-O); δ_H (360 MHz, C^2HCl_3) 0.95 (12H, m., $(CH_3)_2$ x 2), 1.47 (9H, s., $((CH_3)_3C)$, 1.50-1.80 (6H, m., $C\underline{H}_2C\underline{H}$ x 2), 4.06 (1H, app. d. of t., *J* 8.1 and 8 Hz, NHC<u>H</u> Leu1) 4.32 (1H, d.d.d., *J* 8.6, 8 and 8 Hz, NHC<u>H</u> Leu2), 6.96 (1H, d., *J* 8.6 Hz, NH Leu2) and 8.02 (1H, d., *J* 8.1 Hz, NH Leu1); δ_C (90.56 MHz, C^2HCl_3) 23.7 & 24.9 ($(CH_3)_2$ x2), 26.7 (CH x2), 30.2 ($(\underline{C}H_3)_3C$), 42.8 & 43.3 (CH_2 x2), 52.7 (Leu 1 NHCH), 54.8 (Leu 2 NHCH), 82.2 ($(CH_3)_3\underline{C}$), 158.0 (urethane CO) and 174.8 & 177.6 (amide CO); *m/z* (CI) 362 ($[M+NH_4]^+$, 10.6%), 345 ($[M+H]^+$, 100), 306 ($[M+NH_4-(CH_3)_3C]^+$, 28.0), 289 ($[M+H-(CH_3)_3C]^+$, 40.1), 245 ($[M+H-Boc]^+$, 84.6) and 86 ($NH_2CHCH_2CH(CH_3)_2^+$, 45.1).

(2S)-Leucyl-(2S)-leucine trifluoroacetate (28).-

Dipeptide (**27**) (686 mg, 2 mmol) was dissolved in dichloromethane (10 ml) at 0 °C and trifluoroacetic acid (2 ml) added with stirring. After 1.5 h. at 0 °C the solvent was removed under reduced pressure to give a yellow oil which was crystallised from methanol/ light petroleum to give the product as a white solid (406 mg, 84%) of m.p. 250-252 °C (lit.[447] 259-261 °C). $[\alpha]_D^{20}$ -12.1° (c 1 in MeOH) (lit.[447] $[\alpha]_D^{20}$ -13.4° (c 1 in MeOH)); ν_{max} (Nujol) 1745 (acid CO), 1708 (amide CO) and 1480 cm^{-1} (amide II); δ_H (300 MHz, d_6-DMSO) 1.03 (6 H, d., *J* 6.5 Hz, $(CH_3)_2$), 1.05 (6 H, d., *J* 6.5 Hz, $(CH_3)_2$), 1.67 (4H, m., CH_2 x2), 1.75 (2H, m., CH x2), 3.88 (1H, t., *J* 7 Hz, Leu1

NHCH), 4.38 (1H, m., Leu2 NHCH) and 8.83 (1H, d., *J* 7.5 Hz, NH); δ_C (90.56 MHz, d_6-DMSO) 23.78 ($(CH_3)_2$), 24.92 ($(CH_3)_2$), 25.41 & 26.20 (Leu CH's), 42.02 & 43.84 (CH_2), 52.16 & 53.65 (NHCH x2), 173.25 (acid CO) and 176.80 (amide CO).

Methyl N-*tert*-butoxycarbonyl-(2S)-leucyl-(2S)-leucyl-(2S)-leucinate (29).-

Dipeptide (**27**) (688 mg, 2 mmol) was dissolved in dry THF (15 ml), N-methyl morpholine (218 µl, 2 mmol) added and the solution cooled to -15 °C. 2-methylpropylchloroformate (272 µl, 2 mmol) was added and the solution stirred for 1 min. Methyl (2S)-leucinate hydrochloride (**21**) (362 mg, 2 mmol) in DMF (3 ml) containing N-methylmorpholine (218 µl, 2 mmol) was added in one portion. The solution was stirred at -15 °C for 1 min. and at room temperature for 5 min. The hydrochlorides were then filtered off and the THF removed under reduced pressure. Addition of water precipitated a white solid, which was recrystallised from methanol/ water to give the product as a white solid (740 mg, 79%) of m.p. 150-152 °C (lit.[448] 156-159 °C). *m/z* (Found: $(M + H)^+$ 472.3387. $C_{24}H_{46}N_3O_6$ requires 472.3386); $[\alpha]_D^{20}$ -20.3° (c 1 in $CHCl_3$) (lit.[448] $[\alpha]_D^{20}$ -21.7° (c 1 in $CHCl_3$)); ν_{max} (Nujol) 3392 (NH), 1751 (ester CO), 1693 (urethane CO), 1668 (amide CO), 1643 (amide CO), 1541 (amide II) and 1167 cm^{-1} (C-O); δ_H (360 MHz, C^2HCl_3) 0.95 (18 H, m., $(CH_3)_2$ x 3), 1.44 (9H, s., $(CH_3)_3C$), 1.45-1.70 (9H, m., CH_2CH x3), 3.73 (3H, s., OCH_3), 4.10 (1H, m., Leu 3 NHCH), 4.46 (1H, m., Leu 1 NHCH), 4.57 (1H, m., Leu 2 NHCH), 4.76 (1H, d., *J* 7.4 Hz, Leu 2 NH) and 6.63 (2H, m., Leu 1 and 3 NH); δ_C (90.56 MHz, C^2HCl_3) 21.9, 22.1, 22.9 & 23.0 (CH_3 x 6), 24.7 & 24.9 (CH x2), 28.4 ($(CH_3)_3C$), 40.7, 41.5 & 42.0 (CH_2 x3), 50.8 (Leu 1 NHCH), 51.7 (NHCH), 52.4 (CH_3O), 53.2 (NHCH), 80.3 ($(CH_3)_3C$) and 171.6, 172.8 & 173.2 (CO); *m/z* (FAB) 472 ($[M+H]^+$, 30%), 416 ($[M+2H-(CH_3)_3C]^+$, 22.5), 327 (Boc.Leu.Leu.CO^+, 4.4), 227 (H_2NLeu.Leu.CO^+, 8.6), 199 (H_2NLeu.Leu$^+$, 5.5), 146 (H_3NLeuCO_2Me^+, 35.7) and 86 ($H_3NCHCHCH(CH_3)_2^+$, 100).

N-*tert*-Butoxycarbonyl-(2S)-leucyl-(2S)-leucyl-(2S)-leucine (30).-

This was prepared as for (**27**), using (**29**), (471 mg, 1 mmol) and 1 M sodium hydroxide (1.1 ml). Work-up gave the product as a white solid (393 mg, 86%) of m.p. 146-148 °C (lit.[448] 146-147 °C) after recrystallisation from methanol/ water. *m/z* (Found: (*M* + H)$^+$ 457.3152. $C_{23}H_{43}N_3O_6$ requires 457.31516); $[\alpha]_D^{20}$ -28.6 (c 1 in MeOH); ν_{max} (Nujol) 3311 (NH), 3150-2550 (OH), 1724 (acid CO) and 1691, 1669 & 1641 cm^{-1} (amide CO); δ_H (300 MHz, d_6-DMSO) 0.89 (18H, m., $((CH_3)_2$ x3), 1.43 (9H, s., $((CH_3)_3C)$ 1.65 (6H, m., CH_2 x3), 1.72 (3H, m., CH x3), 3.87 (1H, t., *J* 7.3 Hz, Leu1 NHC<u>H</u>), 4.37 (1H, m., Leu3 NHC<u>H</u>), 4.55 (1H, m., Leu2 NHC<u>H</u>), 8.34 (1H, d., *J* 6.4 Hz, NH) and 8.72 (1H, d., *J* 6.8 Hz, NH); δ_C (50.31 MHz, C^2HCl_3) 21.9, 22.1, 22.9 & 23.0 (CH_3 x 6), 23.1 & 24.3 (CH), 28.3 ($(\underline{C}H_3)_3C$), 40.4, 41.5 & 42.0 (CH_2 x3), 50.4, 50.7 & 52.1 (NHCH's), 78.2 ($(CH_3)_3\underline{C}$), 155.6 (urethane CO) and 172.0, 172.2 & 174.1 (amide and acid CO); *m/z* (FAB) 458 (*M*$^+$, 10.9%), 402 ([*M*+2H-$(CH_3)_3C$]$^+$, 12.5), 327 (Boc.Leu.Leu.CO$^+$, 4.4), 227 (H_2NLeu.Leu.CO$^+$, 7.4), 142 (H_3NLeuCO$_2$H$^+$, 6.8) and 86 (H_3NCHCHCH$(CH_3)_2^+$, 100).

(2S)-Leucyl-(2S)-leucyl-(2S)-leucine trifluoroacetate (31).-

This was prepared in an identical manner to (**28**), using (**30**), (365 mg, 0.8 mmol) to give the product as a white solid (328 mg, 92%) of m.p. 203-205 °C. *m/z* (Found: ([*M* + H -CF_3CO_2H)$^+$ 357.2625. $C_{18}H_{35}N_3O_4$ requires 357.2627); $[\alpha]_D^{20}$ -12.4° (c 2 in MeOH); ν_{max} (Nujol) 3274 (NH), 3050 (CH, CH_2), 1730 (acid CO), 1688 & 1676 (amide CO) and 1559 cm^{-1} (amide II); δ_H (300 MHz, d_6-DMSO) 0.89 (18 H, m., $(CH_3)_2$ x3), 1.61 (6H, m., CH_2), 1.74 (3H, m., CH x3), 3.83 (1H, app. t., *J* 7.3 Hz, Leu1 NHC<u>H</u>), 4.36 (1H, m., Leu3 NHC<u>H</u>), 4.55 (1H, m., Leu2 NHC<u>H</u>), 8.32 (1H, d., *J* 7.1 Hz, NH) and 8.72 (1H, d., *J* 6.9 Hz, NH); δ_C (90.56 MHz, C^2HCl_3) 20.9-24.1 (CH_3 x6), 24.91 & 25.24 ($\underline{C}HCH_2$ x3), 40.98, 42.07 & 43.12 ($CH\underline{C}H_2$ x3), 49.94 (Leu 1 NHCH), 50.86 (Leu3 NHCH), 51.10 (Leu2 NHCH), 168.92 (carboxyl CO), 171.26 (Leu 2 CO) and 173.76 (Leu 1 CO); *m/z* (FAB) 358 ([*M*+H]$^+$, 23.9%), 227 (H_2NLeu.Leu.CO$^+$, 19.8), 199 (H_2NLeu.Leu$^+$, 17.0), 131 (H_2NLeu$^+$, 4.6) and 86 (H_2NCHCHCH$(CH_3)_2^+$, 100).

N-*tert*-Butoxycarbonyl-(2S)-leucyl-(2S)-leucyl-(2S)-tyrosine amide (32).-

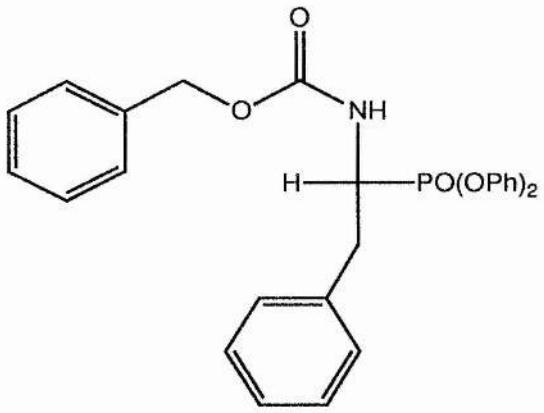
The dipeptide (**27**), (1.72 g, 5 mmol) was dissolved in THF, N-methyl morpholine (550 µl, 5 mmol) added and the solution cooled to -15°C. 2-methylpropylchloroformate (680 µl, 5 mmol) was added with stirring, the mixture stirred for a further 1 min. and (2S)-tyrosine amide (**8**) (905 mg, 5 mmol) in DMF (10 ml) containing N-methyl morpholine (550 µl, 5 mmol) added in one portion. The mixture was stirred at -15 °C for 1 min., then at room temperature for a further 5 min. The hydrochlorides were filtered off and the THF removed under reduced pressure to give an oil. Addition of water (10 ml) to the oil caused the precipitation of a white solid that was recrystallised from ethanol/ water to give the product as a white solid (1.87 g, 74%) of m.p. 189-192 °C (dec.). *m/z* (Found: $(M + H)^+$ 507.3181. $C_{26}H_{43}N_4O_6$ requires 507.3182); $[\alpha]_D^{20}$ +19.3° (c 3 in $CHCl_3$); ν_{max} (Nujol) 3336 (1° amide NH), 3260-3050 (2° amide and urethane NH), 1688 (urethane CO), 1672-76 (2° amide CO), 1657 (1° amide CO), 1557 (amide II) and 1518 cm^{-1} (aromatic); δ_H (300 MHz, d_6-DMSO) 0.93 (12 H, m., $(CH_3)_2$ x2), 1.45 (9H, s., $(CH_3)_3C$), 1.56 (4H, m., $C\underline{H}_2CH$ x2), 1.69 (2H, m., $C\underline{H}CH_2$ x2), 2.72 (1H, d.d.'s, J_{AB} 13.5 Hz, J_{AX} 6.1 Hz, 1 of Tyr CH_2), 2.84 (1H, d.d.'s, J_{AB} 13.5 Hz, J_{BX} 7.5 Hz, 1 of Tyr CH_2), 3.91 (1H, m., Leu 1 NHC$\underline{H}$), 4.13 (1H, m., Leu 2 NHC$\underline{H}$), 4.22 (1H, m., Tyr NHC$\underline{H}$), 6.52 (2H, d., *J* 7.3 Hz, aromatic), 6.96 (2H, d., *J* 7.3 Hz, aromatic), 6.94 (1H, s., 1 of NH_2), 7.01 (1H, s., OH), 7.28 (1H, s., 1 of NH_2), 7.68 (1H, d., *J* 7.1 Hz, NH) and 7.80 (1H, d., *J* 6.2 Hz, NH); δ_C (74.47 MHz, d_6-DMSO) 25.52, 25.66, 26.98 & 27.12 ($(\underline{C}H_3)_2$ x2), 27.91 & 28.22 ($\underline{C}H(CH_3)_2$ x2), 32.14 ($(\underline{C}H_3)_3C$), 40.76 ($ArCH_2$), 44.60 & 44.87 ($CH\underline{C}H_2$ x2), 55.03, 56.70 & 57.77 (NHCH x3), 79.04 ($(CH_3)_3\underline{C}$), 118.86, 131.61 & 134.06 (aromatic), 159.71 (urethane CO) and 175.58, 176.43 & 176.67 (amide CO); *m/z* (FAB) 507 ($[M+H]^+$, 34.6%), 407 ($[M+2H\text{-Boc}]^+$, 8.8), 327 ($[M+H\text{-TyrNH}_2]^+$, 6.9), 136 ($H_3NCHCH_2C_6H_4OH^+$, 35.0), 227 ($H_2N.Leu.Leu.CO^+$, 9.1), 214 ($BocLeuCO^+$, 4.2), 199 ($H_2NLeu.Leu^+$, 6.3), 181 ($TyrNH_3^+$, 34.8) and 86 ($H_3NCHCHCH(CH_3)_2^+$, 100).

(2S)-Leucyl-(2S)-leucyl-(2S)-tyrosyl amide trifluroacetate (33).-

This was prepared exactly as for (**28**), using (**32**) (1 g, 1.98 mmol) and trifluroacetic

acid (2 ml) to give the product as a white solid (700 mg, 87%) of m.p. 207-211°C after recrystallisation from methanol/ light petroleum. $[\alpha]_D^{20}$ +12.9° (c 1 in MeOH); ν_{max} (Nujol) 3321 (1° amide NH), 1674-80 (2° amide CO), 1652 (1° amide CO), 1551 (amide II) and 1519 cm^{-1} (aromatic); δ_H (300 MHz, d_6-DMSO) 0.89 (12H, m., $(CH_3)_2$), 1.14 (2H, m., CH x2), 1.4-1.7 (4H, m., Leu CH_2 x2), 2.76 (1H, d.d.'s, J_{AB} 14.1 Hz, J_{AX} 6.8 Hz, 1 of Tyr CH_2), 2.88 (1H, d.d.'s, J_{AB} 14.1 Hz, J_{BX} 7.7 Hz, 1 of Tyr CH_2), 3.72 (1H, m., Leu1 NHC<u>H</u>), 4.08 (1H, m., Tyr NHC<u>H</u>), 4.37 (1H, m., Leu2 NHC<u>H</u>), 6.62 (2H, d., J 7.6 Hz, aromatic), 6.93 (2H, d., J 7.6 Hz, aromatic), 7.21 (1H, s., OH), 7.73 (1H, d., J 8.4 Hz, NH), 8.05 (2H, br., Leu1 NH_2), 8.48 (1H, d., J 7.6 Hz, NH); δ_C (74.47 MHz, d_6-DMSO) 21.60, 21.77, 22.60, & 23.42† ($(CH_3)_2$ x2), 26.76 & 28.37 ($(CH_3)_3$<u>C</u>H), 43.72, 44.11 & 44.53† (CH_2 x3), 54.55, 55.81 & 56.15 (NHCH), 114.82, 127.64 & 129.92 (aromatic), 158.03 (urethane CO) and 168.74, 170.82 & 172.65 (amide CO).

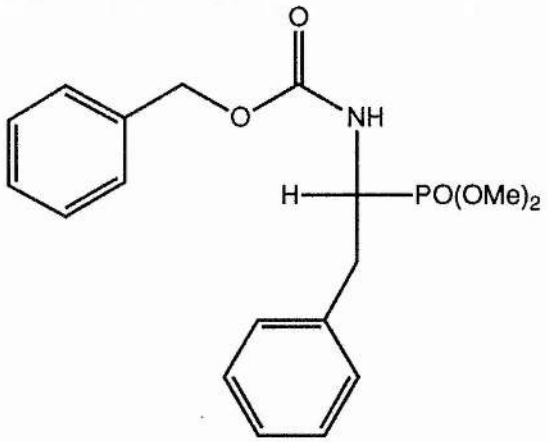
Diphenyl [1-(N-benzyloxycarbonyl)-amino]benzylphosphonate (34).-[332]



Triphenyl phosphite (31.0 g, 0.1 mol), phenylacetaldehyde (18 g, 0.15 mol) and benzyl carbamate (15.1 g, 0.1 mol) were dissolved in glacial acetic acid (15 ml) and stirred at room temperature until the exothermic reaction had ceased (about 1 h.). The mixture was heated to 80 °C for a further 1 h., then the volatile products were removed under reduced pressure on a boiling water bath. The yellow residue was dissolved in methanol (150 ml) and left to crystallise at -10 °C for 3 h. The crystalline product was then collected by filtration, dissolved in the minimum quantity of hot chloroform and a 4 fold excess of methanol added to give the product as a white

solid (14.1 g, 29%) of m.p. 120-2 °C (lit.[332] 119-120 °C). (Found C, 68.75; H 5.44; N, 2.71. Calc. for $C_{28}H_{26}NO_5P$: C, 68.97; H, 5.38; N, 2.87); *m/z* (Found: $(M + H)^+$ 487.1549. $C_{28}H_{26}NO_5P$ requires 487.1549); ν_{max} (Nujol) 3287 (NH), 1720 (urethane CO), 1590 (amide II), 1262 (PO) and 1212 cm^{-1} (P-OPh); δ_H (300 MHz, C^2HCl_3) 3.1 - 3.5 (2H, m., $C\underline{H}_2$-CH), 4.75-4.9 (1H, m., CH), 5.12 (2H, s., $PhCH_2O$), 5.34 (1H, d., *J* 10.5 Hz, NH) and 7.02-7.40 (20 H, m., aromatic); δ_P (121.5 MHz, C^2HCl_3) 17.17 (88%) and 16.74 (12%); δ_C (50.31 MHz, C^2HCl_3) 36.50 ($Ph\underline{C}H_2CH$), 48.22 & 51.35 (d., $PhCH_2\underline{C}H$, J_{PC} 158.1 Hz), 67.7 ($PhCH_2O$), 120.86-121.25 (*para* aromatic), 125.81-130.37 (aromatic), 136.16 & 136.77 (quat. aromatic) and 150.47 (urethane CO); *m/z* (EI) 487 (M^+, 5.4%), 394 ([*M*-OPh]$^+$, 32), 379 ([*M*-$PhCH_2OH$]$^+$, 15.2), 336 ([*M*-$H_2NCOOCH_2Ph$]$^+$, 100), 234 ($PO(OPh)_2^+$, 12.9), 145 ($PhCH_2CHNHCO^+$, 37), 94 (PhOH, 65) and 77 (Ph^+, 14.9).

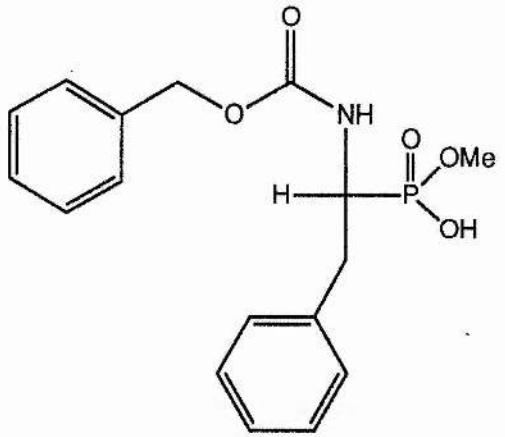
Dimethyl [1-(N-benzyloxycarbonyl)-amino]benzylphosphonate (35).-



The diphenyl compound (**34**) (15.4 g, 31 mmol) was dissolved with warming in dry methanol (200 ml) and sodium (1.47 g, 64 mmol) dissolved in dry methanol (30 ml) added dropwise with stirring. The solution was stirred at room temperature overnight, then the volume reduced to 20 ml on a rotary evaporator and dichloromethane (200 ml) added. The solution was washed with 5% sodium hydrogen carbonate (3 x 80 ml), 0.1 M HCl (80 ml), water (80 ml) and brine (80 ml). The solution was dried ($MgSO_4$) and the solvent removed under reduced pressure to give a yellow oil. Purification by flash chromatography (ethyl acetate) gave a clear oil, which

crystallised on standing. Recrystallisation from ethyl acetate/ light petroleum gave the product as a white solid of m.p. 67-69 °C (7.56 g, 67%). *m/z* (Found: $(M + H)^+$ 364.1314. $C_{18}H_{22}NO_5P$ requires 364.1315); ν_{max} (Nujol) 3224 (NH), 1710 (urethane CO), 1586 (amide II), 1258 (P=O) and 1057 cm^{-1} (P-OMe); δ_H (200 MHz, C^2HCl_3) 2.84-3.3 (2H, m., $PhC\underline{H}_2CH$), 3.62 (3H, d., J_{PH} 10.6 Hz, $P(O)(OCH_3)_2$), 3.68 (3H, d., J_{PH} 10.6 Hz, $P(O)(OCH_3)_2$), 4.40 (1H, d.t., J 7.0 Hz and 8.6 Hz, $PhCH_2C\underline{H}$), 4.97 (2H, s., $PhCH_2O$), 5.81 (1H, d., J 7.1 Hz, NH) and 7.20 (10H, m., Ar); δ_P (121.5 MHz, C^2HCl_3) 26.83 (91%) and 26.21 (9%); δ_C (50.31 MHz, C^2HCl_3) 34.51 ($Ph\underline{C}H_2CH$), 47.44 & 50.59 (d., $PhCH_2\underline{C}H$, J_{PC} 156.4 Hz), 52.82 (d., J_{PC} 27.1 Hz, $-OCH_3$), 65.51 ($PhCH_2O$), 126.79-129.36 (aromatic), 137.48 & 137.75 (quat. aromatic) and 156.14 (urethane CO); *m/z* (CI) 364 ($[M+H]^+$, 100%), 259 ($[M+2H-PhCH_2O]^+$, 12.1), 230 ($[M+H-PhCH_2OCO]^+$, 2.4), 120 ($PhCH_2CHNH_2^+$, 16.6), 108 ($PhCH_2OH$, 11.2) and 91 ($PhCH_2^+$, 13.3).

Methyl hydrogen [1-(N-benzyloxycarbonyl)-amino]-benzylphosphonate (36).-



The dimethyl compound (**35**) (3.63 g, 10 mmol) was dissolved in methanol (30 ml) and 2 M sodium hydroxide (15 ml) added with stirring. The solution was stirred at room temperature for 12 h., the solution diluted with water (60 ml) and washed with dichloromethane (100 ml). The aqueous layer was acidified to pH 2 with 6 M HCl and extracted with dichloromethane (3 x 70 ml). The organic extracts were dried ($MgSO_4$) and evaporated under reduced pressure to give a white solid.

Recrystallisation from ethyl acetate/ light petroleum gave the product as a white solid (2.69 g, 77%) of m.p. 126.5-127.5 °C. *m/z* (Found: (M + H^+) 350.1152. $C_{17}H_{21}NO_5P$ requires 350.1154); ν_{max} (Nujol) 3289 (OH), 2665 and 2320 (P-OH), 1687 (urethane CO), 1546 (amide II), 1219 (P=O), 1041 (P-OH) and 982 cm^{-1} (PO-C); δ_H (200 MHz, C^2HCl_3) 2.93 (2H, m., $PhC\underline{H}_2CH$), 3.62 (3H, d., J_{PH} 10.6 Hz, $P(O)OCH_3$), 4.40 (1H, d.t., J_{NH} 7.1 Hz, J_{CH} 6.3 Hz, $PhCH_2C\underline{H}$), 4.97 (2H, s., $PhCH_2O$), 5.64 (1H, br., OH), 5.81 (1H, d., J 7.1 Hz, NH) and 7.15 (10H, m., Ar); δ_P (121.5 MHz, C^2HCl_3) 22.55 (9.5%) and 23.11 (90.5%); δ_C (50.31 MHz, C^2HCl_3) 38.55 ($Ph\underline{C}H_2CH$), 52.34 & 54.31 (d., J_{PC} 103 Hz, $PhCH_2\underline{C}H$), 56.37 (d., J_{PC} 21.1 Hz, OMe), 69.21 ($PhCH_2O$), 126.71-129.36 (aromatic), 141.34 & 142.40 (quat. aromatic) and 156.17 (urethane CO); *m/z* (EI) 350 ([M + H]$^+$, 5.4%), 320 ([M-OMe+2H]$^+$, 100), 242 ([M-OMe-$PhCH_2O$+2H]$^+$, 18.3) and 91 ($PhCH_2^+$, 49.2).

Diethyl 1-(N-benzyloxycarbonyl)-aminobenzylphosphonate (37).-

This was prepared exactly as for (**35**), using the diphenyl compound (**34**) (2.44 g, 5 mmol) in dry ethanol (30 ml) and sodium (0.286 g, 12.5 mmol) dissolved in dry ethanol (10 ml). Work-up as for (**35**) gave the product as a white solid (1.08 g, 56%). δ_H (200 MHz, C^2HCl_3) 2.84-3.3 (2H, m., $PhC\underline{H}_2CH$), 3.62 (3H, d., J_{PH} 10.6 Hz, $P(O)OCH_3$), 4.40 (1H, app. d.t., J_{NH} 7.1 Hz, J_{CH} 6.4 Hz, $PhCH_2C\underline{H}$), 4.97 (2H, s., $PhCH_2O$), 5.64 (1H, br., OH), 5.81 (1H, d., J 7.1 Hz, NH) and 7.15 (10H, m., Ar); δ_C (50.31 MHz, C^2HCl_3) 16.87 (d., J_{PC} 7.1 Hz, $OCH_2\underline{C}H_3$), 36.36 & 36.43 ($Ph\underline{C}H_2CH$), 47.61 & 50.78 (d., J_{PC} 156.8 Hz, $PhCH_2\underline{C}H$), 62.92-63.51 ($O\underline{C}H_2CH_3$), 67.36 ($Ph\underline{C}H_2O$), 127.22-129.74 (aromatic) and 156.16 (urethane CO).

N-*tert*-Butoxycarbonyl-(2S)-proline (38).-[449]

(2S)-Proline (11.5 g, 0.1 mol) was added with stirring to a 1 M sodium hydroxide solution (100 ml). The suspension was diluted with *tert*-butanol (75 ml) and di-*tert*-butyl dicarbonate (22.3 g, 0.1 mol) in *tert*-butanol (10 ml) added dropwise in the course of 1 h. The reaction was stirred at room temperature for 12 h. after the exothermic reaction had subsided. The reaction mixture was then extracted with light

petroleum (2 x 40 ml) and this organic phase was then extracted with saturated aqueous sodium bicarbonate (3 x 30 ml). The aqueous layers were combined, cooled in an ice bath and acidified to pH 1.5 with 1 M HCl, causing the precipitation of a white solid, which was filtered and dried. The aqueous filtrate was then extracted with ethyl acetate (3 x 50 ml) and the combined organic layers were washed with water (2 x 40 ml), dried ($MgSO_4$) and filtered. The solvent was removed under reduced pressure and the resulting yellowish oil dissolved in light petroleum (15 ml) and kept at -10 °C overnight to give a white solid. The two solids were combined and recrystallised from diethyl ether/ light petroleum to give the product as a white solid (18.9 g, 88%) of m.p. 134-136 °C (lit.[450] 136-7 °C). *m/z* (Found: $(M + H)^+$ 216.1236. $C_{10}H_{17}NO_4$ requires 216.1238); $[\alpha]_D^{20}$ -58° (c 2 in AcOH) (lit.[449] -60.6° (c 2 in AcOH)); ν_{max} (Nujol) 1737 (urethane CO), 1720 (acid CO) and 1554 cm^{-1} (amide II); δ_H (300 MHz, C^2HCl_3) 1.44 & 1.45† (9H, s., $((CH_3)_3C)$, 1.7-2.3 (4H, m., CH_2CH_2), 3.35-3.6 (2H, m., CH_2N), 4.19 (1H, m., CH) and 8.94 (1H, br.s., CO_2H); δ_C (50.31 MHz, C^2HCl_3) 24.13 & 24.78† (C-4), 28.74 & 28.86† ($(\underline{C}H_3)_3C$), 29.42 & 31.3† (C-3), 46.81 & 47.37† (C-5), 59.43 (C-2), 80.84 & 81.55† ($(CH_3)_3\underline{C}$), 154.25 & 156.73† (urethane CO) and 176.59 & 179.46† (acid CO); *m/z* (CI) 216 ($[M+H]^+$, 26.5%), 160 ($[M+2H-(CH_3)_3C]^+$, 59.8), 116 ($[M+2H-(CH_3)_3C-OCO]^+$, 51.5), 70 ($C_4H_9N^+$, 100) and 57 ($(CH_3)_3C^+$, 100).

N-Benzyloxycarbonyl-(2S)-leucine (39).-

To a vigorously stirred solution of (2S)-leucine (1.31 g, 10 mmol) in 5% sodium bicarbonate solution (40 ml) was added benzyloxychloroformate (1.87 g, 11 mmol) dropwise. After 3 hr. the mixture was washed with diethyl ether (2 x 25 ml), acidified to pH 2 with 2 M HCl and extracted with diethyl ether (3 x 30 ml). The ethereal extracts were washed with water (2 x 30 ml), dried ($MgSO_4$) and the solvent removed under reduced pressure to give a thick straw coloured oil (1.835 g, 69%). *m/z* (Found: $(M + H)^+$ 266.1390. $C_{14}H_{19}NO_4$ requires 266.1392); $[\alpha]_D^{20}$ -18° (c 2 in ethanol); δ_H (200 MHz, C^2HCl_3) 0.96 (6H, d., *J* 4.6 Hz, $(CH_3)_2CH$), 1.68 (3H, m., $CHCH_2$), 4.42 (1H, m., $C\underline{H}NH$), 5.12 (2H, s., CH_2OAr), 5.32 (1H, d., *J* 6.8 Hz, NH) and 7.34 (5H, s., aromatic); δ_C (50.31 MHz, C^2HCl_3) 22.08 & 23.36 ($CH(\underline{C}H_3)_2$), 25.23 ($\underline{C}H(CH_3)_2$), 41.89 (CH_2), 52.87 (CHNH), 68.18 (CH_2OAr), 156.78 (urethane

CO) and 178.43 (acid CO); *m/z* (CI) 283 ([*M* + H]$^+$, 32.1%), 266 ([*M* + H]$^+$, 85), 222 ([*M* + H - CO_2]$^+$, 85), 176 ([*M*+H-PhCH_2]$^+$, 18.3), 132 (H_3N-$CH(CO_2H)CH_2CH(CH_3)_2^+$, 23.4) and 91 ($PhCH_2^+$, 59.9).

_p_-Methoxyphenethanol (40).-

p-Methoxyphenylacetic acid (8.4 g, 50 mmol) was dissolved in dry THF (50 ml) and cooled to -20 °C. Triethylamine (5.49 g, 50 mmol) and ethyl chloroformate (5.49 g, 50 mmol) were added with stirring, then the solution was filtered into a solution of sodium borohydride (2.88 g, 75 mmol) in water (60 ml) cooled to 0 °C. The solution was stirred for 3 h. at room temperature, then acidified to pH 2 with 5 M HCl and extracted with diethyl ether (3 x 80 ml). The ethereal extract was dried ($MgSO_4$) and the solvent removed under reduced pressure to give a light yellow oil. Distillation under reduced pressure gave a clear oil that solidified on standing to give a white solid (5.01 g, 66%) identical to that produced by method 2.

Method 2

p-Methoxyphenylacetic acid (8.3 g, 50 mmol) was dissolved in dry diethyl ether (100 ml) and added slowly to lithium aluminium hydride (2.09 g, 55 mmol) suspended in dry diethyl ether (50 ml) at room temperature. The solution was stirred for 15 min., then excess water was cautiously added, followed by 1 M sulphuric acid (25 ml). The layers were separated and the aqueous layer further extracted with diethyl ether (2 x 50 ml). The organic extracts were combined, dried ($MgSO_4$) and the solvent removed under reduced pressure to give a yellow oil. Bulb to bulb distillation under reduced pressure (141-143 °C, 15 mmHg) (lit. [451] 154.5-155 °C, 21 mm Hg)) gave the product as a clear oil that crystallised on standing to give white crystals of m.p. 30-32 °C (5.28 g, 70%). *m/z* (Found: (*M* + H)$^+$ 152.0837. $C_9H_{12}O_2$ requires 152.0837); ν_{max} ($CHCl_3$) 3436 (OH), 3018 (aromatic) and 1613 & 1512 cm^{-1} (aromatic C-H); δ_H (200 MHz, C^2HCl_3) 1.77 (1H, s., OH), 2.82 (2H, t., *J* 6.7 Hz, CH_2Ph), 3.80 (3H, s., OCH_3), 3.82 (2H, t., *J* 6.7 Hz, $C\underline{H}_2OH$), 6.86 (2H, d., *J* 8.7 Hz, aromatic) and 7.15 (2H, d., *J* 8.7 Hz, aromatic); δ_C (50.31 MHz, C^2HCl_3) 38.75 (CH_2Ar), 55.76 (CH_2OH), 64.29 (CH_3O), 114.54 (aromatic, *ortho* to CH_3O), 130.45

(aromatic, *para* to CH_3O), 130.96 (aromatic, *meta* to CH_3O) and 158.76 (aromatic, *ipso* to CH_3O); *m/z* (CI) 152 ($[M + H]^+$, 21.6%), 134 ($CH_3OC_6H_4CHCH_2$, 5.2), 121 ($[CH_3OC_6H_4 + NH_4]^+$, 100) and 91 ($PhCH_2^+$, 14.2).

Bis(triphenylphosphine)tetrahydroboratocopper (41).-[452]

Triphenylphosphine (36 g, 137 mmol) was dissolved in chloroform (200 ml) and cuprous chloride (6.6 g, 66 mmol) added slowly. After the solution became homogeneous a slurry of sodium borohydride (2.54 g, 66 mmol) in 95% ethanol (30 ml) was added slowly and the resulting mixture stirred for 15 min. The mixture was then poured into water (60 ml), the organic layer separated and washed with water (2 x 50 ml), dried ($MgSO_4$) and filtered. The resulting yellow solution was treated with diethyl ether (400 ml) and the resulting precipitate collected and dissolved in chloroform (300 ml). To this solution was added triphenylphosphine (8.34 g, 32 mmol) and sodium borohydride (2.14 g, 55 mmol) in 95% ethanol (25 ml) and the solution stirred for 20 min., after which time ethanol (90 ml) was added. The mixture was filtered and the resulting solution evaporated under reduced pressure until crystals began to form. The solution was cooled and the white precipitate collected, washed with diethyl ether and dried in air to give the product as a white solid of m.p. 175-177 °C (lit.[452] 177 °C).

***p*-Methoxyphenylacetyl chloride (42).**-

p-Methoxyphenylacetic acid (1.66 g, 10 mmol) was dissolved in dry benzene (15 ml), a drop of DMF was added and the mixture cooled to 0 °C. Oxalyl chloride (1.92 g, 22 mmol) was added dropwise, the cooling bath removed and the solution stirred at room temperature for 30 min. The volatiles were then removed to give a brown oil that was distilled under reduced pressure to give the product as a colourless liquid (1.25 g, 68%) that was used immediately. δ_H (300 MHz, C^2HCl_3) 3.81 (3H, s., OCH_3), 4.08 (2H, s., CH_2), 6.89 (2H, d., *J* 7.2 Hz, aromatic) and 7.17 (2H, d., *J* 7.2 Hz, aromatic); δ_C (50.31 MHz, C^2HCl_3) 52.7 (CH_2), 55.8 (OCH_3), 114.9 (aromatic, *ortho* to MeO), 125.8 (aromatic, *para* to MeO), 131.1 (aromatic, *meta* to MeO), 159.9 (*ipso*

to MeO) and 172.8 (CO).

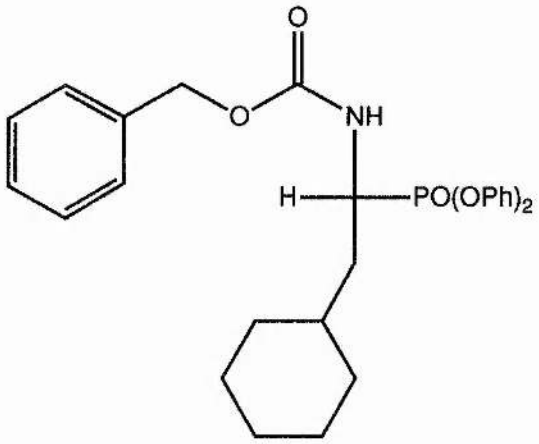
Pyridinium chlorochromate (43).-[453]

Chromium trioxide (10 g, 100 mmol) was added to 6 M HCl (18.4 ml, 110 mmol) rapidly with stirring. After 5 min. the solution was cooled to 0 °C and pyridine (7.91 g, 100 mmol) added slowly over 10 min. Recooling to 0 °C gave the product as an orange solid which was collected by filtration and dried *in vacuo* overnight (17.1 g, 80%).

Cyclohexylacetaldehyde (44).-[340]

2-Cyclohexylethanol (5.08 g, 40 mmol) was dissolved in dry dichloromethane (20 ml) and added in one portion to a suspension of pyridinium chlorochromate (**43**) (12.8 g, 60 mmol) in dry dichloromethane (100 ml). The mixture was stirred for 2 h. at room temperature, then dry diethyl ether (100 ml) added and the solution decanted. The residue was washed with dry diethyl ether (3 x 50 ml), the organics combined and filtered through a pad of Florisil. Evaporation of the solvent gave a light green oil that was purified by distillation under reduced pressure (68-70 °C, 25 mm Hg) to give the product as a colourless oil (3.3 g, 67%). ν_{max} (thin film) 3021, 2927 & 2855 (C-H) and 1721 cm^{-1} (aldehyde CO); δ_H (200 MHz, C^2HCl_3) 0.9-2.0 (10H, m., cyclohexyl CH_2), 2.31 (1H, m., C<u>H</u>-CH_2), 4.14 (1H, m., C<u>H</u>$_2$-CHO) and 9.78 (1H, t., *J* 2.4 Hz, CHO); δ_C (50.31 MHz, C^2HCl_3) 26.51-26.79 (cyclohexyl CH_2 x4), 33.10-33.72 (cyclohexyl CH_2 x2, CH), 51.83 (<u>C</u>H_2-CHO) and 203.45 (CHO).

Diphenyl[1-(N-benzyloxycarbonyl)-amino]cyclohexylmethylphosphonate (45).-



This was synthesised exactly as for the benzylphosphonate analogue (**34**), using aldehyde (**44**) (20 mmol, 2.5 g). The product, after recrystallisation from chloroform/ methanol, was obtained as a white solid of m.p.136-138 °C (4.7 g, 48%). *m/z* (Found: $(M+H)^+$ 494.2095. $C_{28}H_{33}NO_5P$ requires 494.2098); ν_{max} (Nujol) 3273 (NH), 1718 (urethane CO), 1543 (amide II), 1244 (P=O), 954 (P-O) and 763 cm^{-1} (aromatic); δ_H (200 MHz, C^2HCl_3) 0.7-1.9 (11H, m., cyclohexyl), 4.39 (1H, m., CH-P), 5.13 (2H, s., $PhCH_2O$), 7.07-7.15 (15H, m., aromatic) and 8.04 (1H, d., *J* 8.4 Hz, NH); δ_P (121.49 MHz, C^2HCl_3) 17.92; δ_C (75.47 MHz, d_6-DMSO) 25.2-25.82 (cyclohexyl CH_2 x5), 37.17 (d., J_{PC} 10.9 Hz, cyclohexyl CH), 39.50 (d., J_{PC} 46.8 Hz, $CH\underline{C}H_2CHP$), 44.75 (d., J_{PC} 158.6 Hz, CHP), 69.75 (benzyl CH_2), 124.3-131.12 (aromatic), 136.91 (quat. aromatic) and 160.22 (urethane CO); *m/z* (CI) 494 ($[M+H]^+$, 100%), 400 ($[M-OPh]^+$, 15.5), 360 ($[M+2H-Cbz]^+$, 1.2), 260 ($[M-PO(OPh)_2]^+$, 9.6), 126 ($H_2NCH_2C_6H_{11}^+$, 39.4) and 91 ($PhCH_2^+$, 70.4).

Dimethyl 1-(N-benzyloxycarbonyl)-amino cyclohexylmethylphosphonate (46).-

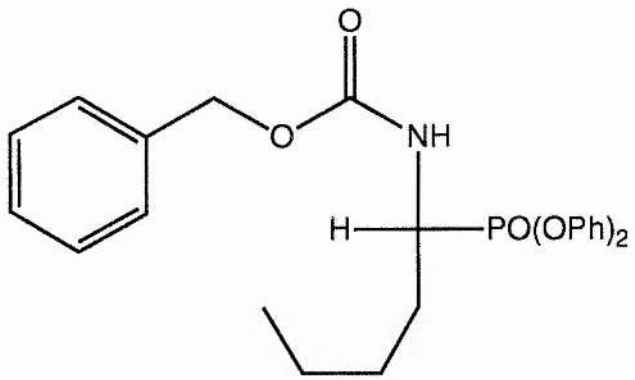
This was prepared exactly as for the benzyl analogue (**35**), using diphenyl ester (**45**) (4.69 g, 10 mmol) in dry methanol (75 ml) and sodium (0.506 g, 22 mmol) in dry

methanol (25 ml). The same work-up gave the product as a white solid of m.p. 83.5-84.5 °C after recrystallisation from chloroform/ light petroleum (2.59 g, 68%). *m/z* (Found: (*M*+ H)$^+$ 369.1701. $C_{18}H_{28}NO_5P$ requires 369.1698); ν_{max} (Nujol) 3271 (NH), 3062 (aromatic CH), 1714 (urethane CO), 1548 (amide II), 1247 (P=O) and 1223 cm^{-1} (P-O); δ_H (300 MHz, C^2HCl_3) 0.7-1.75 (10H, m., CH_2 x5), 1.83 (1H, m., CH), 3.67 (3H, d., J_{PH} 10.9 Hz, OCH_3), 4.20 (1H, m., POCH), 5.10 (2H, s., benzyl), 5.15 (1H, d., J_{NH} 9.6 Hz, NH) and 7.29 (5H, s., aromatic); δ_P (121.49 MHz, C^2HCl_3) 27.82; δ_C (75.4 MHz, C^2HCl_3) 26.7-27.1 (cyclohexyl CH_2), 32.42, 34.74 & 37.61† (obscured d., CH$\underline{C}H_2$CH), 34.48 (d., J_{PC} 13.1 Hz, cyclohexyl CH), 46.7 (d., J_{PC} 154.7 Hz, P(O)CH), 53.9 (d., J_{PC} 28.2 Hz, OCH_3), 68.1 (CH_2Ph), 128.7-129.22 (aromatic), 137.21 (quat. aromatic) and 156.72 (urethane CO); *m/z* (EI) 370 ([*M*+H]$^+$, 5.4%), 260 ([*M*-PO(OMe)$_2$]$^+$, 65.4), 110 ([PO(OMe)$_2$+ H]$^+$, 68.8) and 91 ($PhCH_2^+$, 100).

Methyl hydrogen [1-(N-benzyloxycarbonyl)-amino] cyclohexyl methylphosphonate (47).-

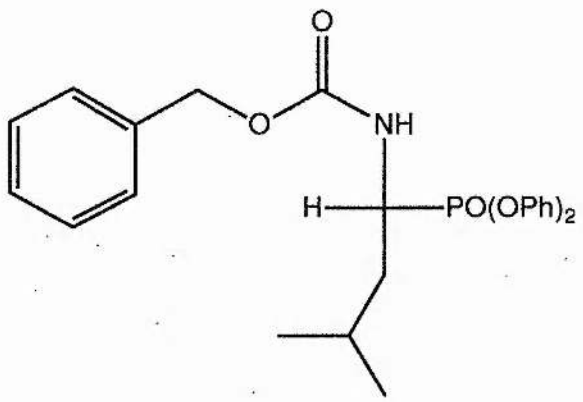
This was synthesised exactly as for the benzylphosphonate analogue (**36**), using compound (**46**) (3.69 g, 10 mmol) in methanol (40 ml) with 2 M sodium hydroxide (15 ml). Recrystallisation from chloroform/ light petroleum gave the product as a white solid of m.p. 126-128 °C (2.76 g, 77%). *m/z* (Found: (*M* + H)$^+$ 356.1628. $C_{17}H_{27}NO_5P$ requires 356.1626); ν_{max} (Nujol) 3305 (OH), 2662 and 2350 (P-OH), 1686 (urethane CO), 1546 (amide II), 1216 (P=O), 1041 (P-OH) and 982 cm^{-1} (PO-C); δ_H (200 MHz, C^2HCl_3) 0.7-1.7 (10H, m., CH_2 x5), 1.83 (1H, m., CH), 3.71 (3H, d., J_{PH} 12.9 Hz, OCH_3), 4.21 (1H, m., CHPO), 5.22 (2H, s., $PhCH_2O$), 6.82 (1H, br., NH), 7.31 (5H, s., aromatic) and 9.88 (1H, br.s., P(O)OH); δ_P (121.49 MHz, C^2HCl_3) 27.92 & 28.69 (diastereomers, 1:5.2); δ_C (75.47 MHz, C^2HCl_3) 26.78-27.15† (cyclohexyl CH_2), 32.82, 35.11 & 37.42† (obscured d., cyclohexyl CH_2), 34.74 (cyclohexyl CH), 45.16 (d., J_{PC} 154 Hz, CH-PO) 53.78 (d., J_{PC} 27.2 Hz, OCH_3), 67.41 (benzyl CH_2), 111.4-111.95 (aromatic), 134.89 & 134.85† (quat. aromatic) and 156.47 (urethane CO); *m/z* (CI) 356 ([*M* + H]$^+$, 100).

Diphenyl 1-(N-Benzyloxycarbonyl)-amino-*n*-butylphosphonate (48).-



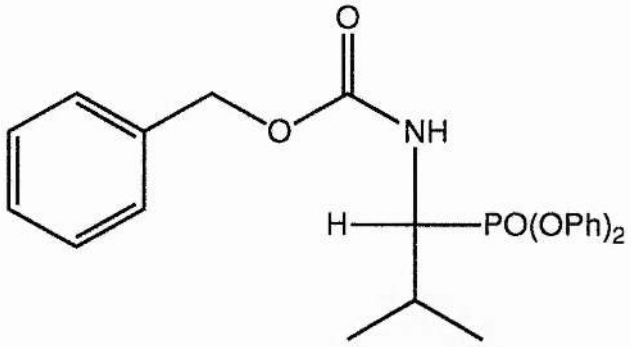
This was prepared exactly as for the benzyl analogue (**34**), using pentanal (12.9 g, 150 mmol). The usual work-up gave the product as a white solid (23.8 g, 53%) of m.p. 102-3 °C. (Found: C, 66.25; H, 6.39; N, 3.12. Calc. for $C_{25}H_{28}NO_5P$: C, 66.20; H, 6.18; N, 3.09%); *m/z* (Found: $(M + H)^+$ 453.1701. $C_{25}H_{28}NO_5P$ requires 453.1698); ν_{max} (Nujol) 3267 (NH), 1710 (urethane CO), 1589 (amide II), 1253 (P=O), 950 (P-O) and 759 cm^{-1} (aromatic); δ_H (300 MHz, C^2HCl_3) 0.88 (3H, t., *J* 7.4 Hz, CH_3), 1.1-1.5 (4H, m., $C\underline{H}_2$-$C\underline{H}_2$), 1.71 (1H, m., 1 of $C\underline{H}_2$-CH-P), 1.97 (1H, m., 1 of $C\underline{H}_2$-CH-P), 4.44 (1H, m., CH_2-C<u>H</u>-P), 5.04 (1H, AB d., *J* 11.8 Hz, 1 of $PhCH_2O$), 5.10 (1H, AB d., *J* 11.8 Hz, 1 of $PhCH_2O$), 5.52 (1H, br., NH) and 6.92-7.31 (15H, m., aromatic); δ_P (121.49 MHz, C^2HCl_3) 18.23 (93%) and 17.92 (7%); δ_C (75.47 MHz, C^2HCl_3) 13.62 (CH_3), 21.95 (C-4 CH_2), 27.50 (d., J_{PC} 2.9 Hz, C-3 CH_2), 29.44 (d., J_{PC} 13.3 Hz, C-2 CH_2), 47.52 (d., J_{PC} 157.7 Hz, C-1 CH), 66.91 and 67.21 ($PhCH_2O$), 120.14-129.88 (aromatic), 136.0 (quat. aromatic), 149.82-150.18 (quat. aromatic ArC-O) and 155.82 & 155.90 (urethane CO); *m/z* (CI) 454 ($[M+H]^+$, 100%), 363 ($[M-PhCH_2]^+$, 25.3), 346 ($[M-PhCH_2OH]^+$, 41.5), 235 $(PO(OPh)_2^+$, 4.2), 220 ($CbzNHCH(CH_2)_3CH_3^+$, 6.3) and 108 ($PhCH_2OH_2^+$, 40.8).

Diphenyl 1-(N-benzyloxycarbonyl)-amino-2-methylpropylphosphonate (49).-



This was prepared exactly as for the benzyl analogue (**34**), using 3-methyl butanal (12.9 g, 150 mmol). The usual work-up gave the product as a white solid (20.1 g, 44.3%) of m.p. 123-4 °C (lit.[332] 122-3 °C). (Found: C, 66.45; H, 6.45; N, 3.09. Calc. for $C_{25}H_{28}NO_5P$: C, 66.20; H, 6.18; N, 3.09%); *m/z* (Found: $(M + H)^+$ 454.1776. $C_{25}H_{29}NO_5P$ requires 454.1778). ν_{max} (Nujol) 3269 (NH), 1713 (urethane CO), 1592 (amide II), 1270 (P=O), 946 (P-O) and 756 cm^{-1} (aromatic); δ_H (300 MHz, C^2HCl_3) 0.91 (6H, d., *J* 5.4Hz, $(CH_3)_2$), 1.62-1.74 (3H, m., CH_2CH), 4.51 (1H, m., CH-PO), 5.12 (2H, s., $PhCH_2O$), 5.19 (1H, obscured d., NH) and 6.87-7.38 (15H, m., aromatic); δ_P (121.49 MHz, C^2HCl_3) 18.23 (14.4%) and 18.57 (85.6%); δ_C (75.47 MHz, C^2HCl_3) 21.62 (CH_3), 23.71 (CH_3), 24.80 & 25.03† (($CH_3)_2\underline{C}H$), 39.02 & 39.98† (d., J_{PC} 48.2 Hz, $(CH_3)_2CH\underline{C}H_2$), 47.43 (d., J_{PC} 158.4 Hz, CH-P), 67.67 & 68.03 ($\underline{C}H_2OAr$), 120.87-130.31 (aromatic), 136.7 (quat. aromatic Ar$\underline{C}$-C), 150.6-150.97 (quat. aromatic Ar$\underline{C}$-O) and 154.45 & 156.38† (urethane CO); *m/z* (CI) 454 ($[M+H]^+$, 100%), 363 ($[M-PhCH_2]^+$, 12.6), 346 ($[M-PhCH_2OH]^+$, 10.8), 320 ($[M+H-Cbz]^+$, 2.1), 220 ($CbzNHCHCH_2CH(CH_3)_2^+$, 6.6) and 108 ($PhCH_2OH_2^+$, 40.8).

Diphenyl 1-(N-benzyloxycarbonyl)-amino-*iso*propylphosphonate (50).-



This was prepared exactly as for the benzyl analogue (**34**), using 2-methyl propanal (10.8 g, 150 mmol). The usual work-up gave the product as a white solid (24.8 g, 56.5%) of m.p. 108-9 °C (lit.[332] 104-5 °C). (Found: C, 65.7; H, 5.92; N, 3.23. Calc. for $C_{24}H_{26}NO_5P$: C, 65.6; H, 5.92; N, 3.19); *m/z* (Found: $(M + H)^+$ 439.1546. $C_{25}H_{28}NO_5P$ requires 439.1548); ν_{max} (Nujol) 3305 (NH), 1716 (urethane), 1591 (amide II), 1245 (P=O), 946 (P-O) and 746 cm^{-1} (aromatic); δ_H (200 MHz, C^2HCl_3) 1.14 (6H, d., *J* 7.3 Hz, $(C\underline{H}_3)_2CH$), 2.51 (1H, m, $(CH_3)_2C\underline{H}$), 4.44 (1H, d.d.d., J_{CH} 4.1, J_{NH} 10.9, J_{PH} 19.1 Hz, CH-PO), 5.13 (1H, AB d., *J* 11.8 Hz, 1 of $PhCH_2O$), 5.16 (1H, AB d., *J* 11.8 Hz, 1 of $PhCH_2O$), 5.29 (1H, d., *J* 10.9 Hz, NH) and 7.03-7.43 (15H, m., aromatic); δ_P (121.49 MHz, C^2HCl_3) 17.86 (18.6%) and 18.21 (81.4%); δ_C (75.47 MHz, C^2HCl_3) 17.55 & 17.60† ($(\underline{C}H_3)_2CH$), 20.1 & 20.28† ($(\underline{C}H_3)_2CH$), 29.12 (d., J_{PC} 8.1 Hz, $(CH_3)_2\underline{C}H$), 52.59 (d., J_{PC} 154.2 Hz, CH-P), 67.24 & 67.45† ($PhCH_2O$), 120.25-129.46 (aromatic), 135.96 (quat. aromatic Ar$\underline{C}$-C), 149.8-150.15 (quat. aromatic, ArC-O-P) and 156.05 & 156.14 (urethane CO); *m/z* (CI) 440 ($[M+H]^+$, 100%), 349 ($[M-PhCH_2]^+$, 12), 206 ($[M-PO(OPh)_2]^+$, 9.3) and 108 ($PhCH_2OH_2^+$, 45.5).

Dimethyl 1-(N-benzyloxycarbonyl)-amino*n*-butylphosphonate (51).-
Method 1

This was prepared exactly as for the benzylphosphonate analogue (**34**), using (**48**) (4.53 g, 10 mmol). The product was isolated as a clear oil after flash

chromatography, identical in all respects to that obtained from method 2, (2.86 g, 79%) (ethyl acetate) that proved refractory to crystallisation.

Method 2[334]

The diphenyl phosphonate (**48**) (4.53 g, 10 mmol) was dissolved in methanol (30 ml) and potassium difluoride dihydrate (8.0 g) and 18-crown-6 (20 mg) added. The mixture was heated to boiling for 5 min. then stirred at room temperature overnight, then the solvent was removed under reduced pressure. The resulting solid was suspended in ethyl acetate (40 ml), washed with water (2 x 20 ml), 5% sodium bicarbonate solution (3 x 20 ml) and brine (20 ml). The organic layer was dried ($MgSO_4$) and the solvent removed under reduced pressure to give a light yellow oil. Flash column chromatography (ethyl acetate) gave a clear oil (2.11 g, 64%) that could not be crystallised. ν_{max} (Nujol) 3327 (NH), 1721 (urethane CO), 1586 (amide II), 1243 (P=O), 941 (P-O) and 749 cm^{-1} (aromatic); δ_H (200 MHz, C^2HCl_3) 0.89 (3H, t., *J* 6.9 Hz, $C\underline{H}_3CH_2$), 1.12-1.63 (4H, m., $C\underline{H}_2C\underline{H}_2$), 1.81 (2H, m., $POCHC\underline{H}_2$), 3.61 (3H, d., J_{PH} 10.5 Hz, 1 of $PO(OCH_3)_2$), 3.64 (3H, d., J_{PH} 10.5 Hz, 1 of $PO(OCH_3)_2$), 4.09 (1H, m., $POC\underline{H}CH_2$), 5.01 (1H, AB d., *J* 12.3 Hz, 1 of $PhCH_2O$) and 5.08 (1H, AB d., *J* 12.3 Hz, 1 of $PhCH_2O$), 5.82 (1H, d., *J* 9.7 Hz, NH) and 7.34 (5H, s., aromatic); δ_P (121.49 MHz, C^2HCl_3) 27.06; δ_C (50.31 MHz, C^2HCl_3) 14.28 (CH_3), 22.59 ($CH_3\underline{C}H_2$), 28.32 (d., J_{PC} 12.8 Hz, $\underline{C}H_2CH$), 29.58 (d., J_{PC} 2.8 Hz, $CH_3CH_2\underline{C}H_2$), 49.20 and 46.09 (d., J_{PC} 156.4 Hz, CHPO), 53.28 (d., J_{PC} 6.7 Hz, 1 of $PO(OCH_3)_2$), 53.61 (d., J_{PC} 6.7 Hz, 1 of $PO(OCH_3)_2$), 128.36-128.86 (aromatic), 136.90 (quat. aromatic) and 156.80 (urethane CO).

Dimethyl 1-(N-Benzyloxycarbonyl)-amino-2-methylpropylphosphonate (52).-Method 1

This was synthesised exactly as for the benzyl analogue (**35**) using (**49**) (4.53 g, 10 mmol), giving the product as a clear oil (2.59 g, 71%) after flash chromatography (ethyl acetate). The oil could not be crystallised.

Method 2[331]

To a stirred suspension of lead tetraacetate (4.48 g, 10.1 mmol) in dry DMF (10 ml) at 0 °C was added dropwise a solution of N-benzyloxycarbonyl-(2S)-leucine (**39**) (2.65 g, 10 mmol) in dry DMF (10 ml). After 30 min., cooling was removed and the reaction stirred for an additional 3.5 h., and then saturated sodium carbonate solution (75 ml) was added. The mixture was extracted with ethyl acetate (4 x 60 ml) and the organic extracts washed with saturated sodium carbonate solution (100 ml) and then brine (60 ml), dried ($MgSO_4$) and evaporated under reduced pressure to give the crude product as a yellowish oil. A solution of this oil in dry dichloromethane (50 ml) was then added to trimethyl phosphite (1.49 g, 11 mmol) and the mixture cooled to -78 °C under an argon atmosphere. Titanium tetrachloride (2.09 g, 11 mmol) in dry dichloromethane (10 ml) was added dropwise over 10 min. and the resulting yellow solution allowed to warm to room temperature overnight. Sodium carbonate decahydrate (5 g, 17.5 mmol) suspended in dichloromethane (60 ml) was then added and the mixture stirred for a further 30 min., filtered, dried ($MgSO_4$) and evaporated under reduced pressure to give a yellow oil. Purification by flash column chromatography (ethyl acetate) gave the product as a clear oil that crystallised on standing to give white crystals (1.71 g, 54 %) of m.p. 44-6 °C. ν_{max} (Nujol) 3321 (NH), 1728 (urethane CO), 1584 (amide II), 1247 (P=O), 941 (P-O) and 744 cm^{-1} (aromatic); δ_H (200 MHz, C^2HCl_3) 0.90 (6H, d., *J* 6.6 Hz, $(C\underline{H}_3)_2CH$), 1.32-1.78 (3H, m., $-C\underline{H}_2C\underline{H}-$), 3.61 (3H, d., J_{PH} 9.8 Hz, 1 of $PO(OC\underline{H}_3)_2$) 3.65 (3H, d., J_{PH} 9.8 Hz, 1 of $PO(OC\underline{H}_3)_2$), 4.14 (1H, m, $POC\underline{H}CH_2$), 5.06 (1H, AB d., *J* 12.3 Hz, 1 of $PhCH_2O$), 5.09 (1H, AB d., *J* 12.3 Hz, 1 of $PhCH_2O$), 5.15 (1H, d., *J* 10.8 Hz, NH) and 7.34 (5H, s., aromatic); δ_P (121.49 MHz, C^2HCl_3) 27.06; δ_C (75.47 MHz, C^2HCl_3) 21.73 (CH_3), 23.86 (CH_3), 24.95 (d., J_{PC} 11.6 Hz, $(CH_3)_2\underline{C}H$), 39.58 (d., J_{PC} 48.1 Hz, $(CH_3)_2CH\underline{C}H_2$), 46.81 (d., J_{PC} 158.4 Hz, CH-PO), 53.75 (d., J_{PC} 7.2 Hz, 1 of $PO(OCH_3)_2$), 53.98 (d., J_{PC} 7.2 Hz, 1 of $PO(OCH_3)_2$), 66.54 (OCH_2Ar), 125.36-128.86 (aromatic), 136.90 (quat. aromatic) and 156.60 (urethane CO).

Dimethyl 1-(N-benzyloxycarbonyl)-amino*iso*propylphosphonate (53).-

This was prepared exactly as for the benzyl analogue (**35**), using (4.39 g, 10 mmol) of (**50**). Crystallisation of the clear oil (1.96 g, 62%) obtained from flash chromatography (ethyl acetate) proved impossible. ν_{max} (Nujol) 3342 (NH), 1721 (urethane CO), 1580 (amide II), 1240 (PO), 941 (P-O) and 752 cm^{-1} (aromatic); δ_H (200 MHz, C^2HCl_3) 0.97 (6H, d., $(C\underline{H}_3)_2CH$), 2.20 (1H, m., $(CH_3)_2C\underline{H}$), 3.71 (3H, d., J_{PH} 6.4 Hz, $PO(OCH_3)_2$), 3.74 (3H, d., J_{PH} 6.4 Hz, $PO(OCH_3)_2$), 4.11 (1H, m., CHPO), 5.12 (2H, s., $PhCH_2O$) and 7.32 (5H, s., aromatic); δ_P (121.49 MHz, C^2HCl_3) 28.14.

Methyl hydrogen 1-(N-benzyloxycarbonyl)-amino*n*-butylphosphonate (54).-

This was produced exactly as for the benzyl analogue (**36**), using dimethyl phosphonate (**51**) (3.29 g, 10 mmol). The resulting solid was recrystallised from chloroform/ light petroleum to give the product as a white powder (2.58 g, 82%) of m.p. 125-127 °C. *m/z* (Found: $(M + H)^+$ 316.1310. $C_{14}H_{23}NO_5P$ requires 316.1308). ν_{max} (Nujol) 3302 (NH), 3140-2810 (br., OH), 1718 (urethane CO), 1585 (amide II), 1240 (P=O), 931 (P-O) and 745 cm^{-1} (aromatic); δ_H (300 MHz, C^2HCl_3) 0.88 (3H, t., *J* 6.9 Hz, CH_3), 1.14-1.62 (4H, m., CH_2CH_2), 1.71 (1H, m., 1 of $C\underline{H}_2$-CHP), 1.97 (1H, m., 1 of $C\underline{H}_2$-CH-P), 3.70 and 3.71 (3H, d., J_{PH} 10.8 Hz, diastereomers of CH_3O-PO), 4.09 (1H, m., CH_2-C<u>H</u>P), 5.17 (2H, s., $PhCH_2O$), 5.19 (1H, d., obscured, NH), 7.31 (5H, m., aromatic) and 10.83 (1H, br., PO-OH); δ_P (121.49 MHz, C^2HCl_3) 28.11 (76.2%) and 27.39 (23.8%); δ_C (75.47 MHz, C^2HCl_3) 14.94 (CH_3), 26.46 ($CH_3\underline{C}H_2$), 27.94 (d., J_{PC} 12.1 Hz, $\underline{C}H_2CH$), 29.02 (d., J_{PC} 2.8 Hz, $CH_3CH_2\underline{C}H_2$), 47.42 & 50.53 (d., J_{PC} 156.4 Hz, CHPO), 55.28 (d., J_{PC} 8.6 Hz, $PO(OCH_3)_2$), 67.94 (CH_2O), 127.36-130.32 (aromatic), 137.87 (quat. aromatic) and 156.67 (urethane CO); *m/z* (CI) 316 ($[M + H]^+$, 29.5%), 298 ($[M\text{-}CH_3(CH_2)_3]^+$, 4) and 108 ($PhCH_2NH_3^+$, 21).

Methyl hydrogen 1-(N-benzyloxycarbonyl)-amino-2-methylpropyl phosphonate (55).-

This was produced exactly as for the benzyl analogue (**36**), using dimethyl phosphonate (**53**) (3.29 g, 10 mmol), giving the product, after recrystallisation from chloroform/ light petroleum, as a white solid of m.p. 119-121 °C (2.24 g, 71%). *m/z* (Found: $(M + H)^+$ 316.1309. $C_{14}H_{23}NO_5P$ requires 316.1308). ν_{max} (Nujol) 3302 (NH), 3140-2850 (br., OH), 1718 (urethane CO), 1585 (amide II), 1240 (P=O), 931 (P-O) and 745 cm^{-1} (aromatic); δ_H (300 MHz, C^2HCl_3) 0.91 (6H, d., *J* 6.5Hz, $CH_2CH(C\underline{H}_3)_2$), 1.58 (2H, m, $C\underline{H}_2CH(CH_3)_2$), 1.74 (1H, m., $CH_2C\underline{H}(CH_3)_2$), 3.71 (3H, d., J_{PH} 10.8 Hz, CH_3O-P), 4.19 (1H, m., CH-PO), 5.09 (1H, obscured d., NH), 5.12 (2H, s., $PhCH_2O$), 7.34 (5H, m., aromatic) and 10.83 (1H, br.s., PO-OH); δ_P (121.49 MHz, C^2HCl_3) 27.62 (18%) and 28.32 (82%); δ_C (75.47 MHz, C^2HCl_3) 20.81-92† ($(\underline{C}H_3)_2CH$), 24.21 (d., J_{PC} 13.1 Hz, $(CH_3)_2\underline{C}H$), 37.85 (d., J_{PC} 34.2 Hz, $\underline{C}H_2CH$), 43.97 & 47.12 (d., J_{PC} 158.5 Hz, PO-CH), 52.08 (d., J_{PC} 10.4 Hz, P-OCH_3), 67.03 & 67.50 ($PhCH_2O$), 127.8-128.4 (aromatic), 136.12 (quat. aromatic) and 155.27 & 156.06 (urethane CO); *m/z* (CI) 333 ($[M + NH_4]^+$, 18%), 316 ($[M + H]^+$, 100) and 91 ($PhCH_2^+$, 58).

Methyl hydrogen 1-(N-benzyloxycarbonyl)-amino*iso*propylphosphonate (56).-

This was prepared exactly as for the benzyl analogue (**36**), using dimethyl phosphonate (**53**), (3.15 g, 10 mmol), giving the product, after recrystallisation from chloroform/ light petroleum, as a white solid of m.p. 102.5-105 °C (1.96 g, 65%). *m/z* (Found: $(M + H)^+$ 302.1160. $C_{13}H_{21}NO_5P$ requires 302.1158). ν_{max} (Nujol) 3302 (NH), 3140-2410 (PO-OH), 1718 (urethane CO), 1585 (amide II), 1240 (P=O), 931 (P-O) and 745 cm^{-1} (aromatic); δ_H (300 MHz, C^2HCl_3) 0.97 & 0.98 (6H, t., *J* 6.9 Hz, diastereomers of $(C\underline{H}_3)_2CH$), ≈ 6:1), 2.20 (1H, m., $(CH_3)_2C\underline{H}$)), 3.68 (3H, d., J_{PH} 10.9 Hz, PO-OCH_3), 4.07 (1H, d.d.d., J_{PH} 19.0 Hz, J_{NH} 10.5 Hz, J_{CH} 4.3 Hz, PO-CH), 5.12 (2H, s., $PhCH_2O$), 5.22 (1H, d., *J* 10.5 Hz, NH), 7.37 (5H, s., aromatic) and 10.83 (1H, br.s., PO-OH); δ_P (121.49 MHz, C^2HCl_3) 27.03 (14%) and 27.78 (86%); δ_C (75.47

MHz, C^2HCl_3) 17.48 (d., J_{PC} 2.5 Hz, $(\underline{C}H_3)_2CH$, 1 diastereomer), 20.19 (d., J_{PC} 5.7 Hz, $(\underline{C}H_3)_2CH$, 1 diastereomer), 28.62 (d., J_{PC} 10.6 Hz, $(CH_3)_2\underline{C}H$), 51.16 & 53.22 (d., PO-CH, J_{PC} 155.5 Hz), 52.42 & 53.15 (d., J_{PC} 15.8 Hz, PO-OCH_3, diastereomers), 67.15 & 67.43 ($PhCH_2O$, diastereomers), 127.77-128.38 (aromatic), 136.07 (quat. aromatic) and 156.27 & 156.36 (urethane CO); *m/z* (CI) 302 ([*M* + H]$^+$, 43%), 284 ([*M*-OH]$^+$, 28.1) and 91 ($PhCH_2^+$, 56).

Dimethyl phthalimidomethylphosphonate (57).- Method 1

To a 50% dispersion of sodium hydride in mineral oil (1.58 g, 33 mmol) in dry THF (10 ml) under nitrogen was added dimethyl phosphite (3.3 g, 30 mmol) in dry THF (30 ml). The mixture was stirred at room temperature for 45 min., until effervescence had ceased, and allowed to settle. The solution was then added dropwise to N-bromomethylphthalimide (7.2 g, 30 mmol) in dry toluene (50 ml) under nitrogen at 0 °C. The reaction was refluxed for 8 h., filtered and the pad washed with ethyl acetate. The THF was removed under reduced pressure and the solution washed with water (40 ml) and brine (40 ml), dried ($MgSO_4$) and evaporated under reduced pressure to give a white solid. Crystallisation from ethyl acetate/ light petroleum gave the product as white amorphous crystals (6.72 g, 83%) of m.p. 103-6 °C. *m/z* (Found: (*M* + H)$^+$ 270.0530. $C_{11}H_{13}NO_5P$ requires 270.0528); ν_{max} (Nujol) 1718 (imide CO), 1254 (P=O), 1020 (P-OMe) and 835 cm^{-1} (aromatic); δ_H (200 MHz, C^2HCl_3) 3.87 (6H, d., *J* 11.3 Hz, $P(OCH_3)_2$), 4.13 (2H, d., J_{PH} 9.8 Hz, $C\underline{H}_2$-$PO(OMe)_2$) and 7.71-8.0 (4H, m., aromatic); δ_P (121.49 MHz, C^2HCl_3) 22.50; δ_C (50.31 MHz, C^2HCl_3) 29.78 & 32.81 (d., J_{PC} 151 Hz, CH_2-P), 52.12 (d., J_{PC} 5.7 Hz, $PO(OCH_3)_2$), 122.21 (aromatic), 130.82 (quat. aromatic), 133.82 (aromatic) and 165.84 (CO); *m/z* (EI) 269 (*M*$^+$, 18.8%), 160 ([*M*-$PO(OMe)_2$]$^+$, 100), 133 ([$C_6H_4(CO)_2$ + H]$^+$, 12.5) and 76 ($C_6H_4^+$, 18.8).

Method 2[348]

N-bromomethylphthalimide (2.4 g, 10 mmol) was dissolved in 5 ml xylene and trimethyl phosphite (1.36 g, 11 mmol) added. The reaction flask was swept with

argon and heated slowly to near reflux, when the mixture reacted exothermically. Heating was discontinued until the exotherm had subsided, then the mixture was refluxed for 5 h. Most of the solvent was then removed under reduced pressure and the remainder allowed to cool overnight. The resulting yellow solid was collected, heated in dry diethyl ether (15 ml) for 3 h., filtered and evaporated to give the crude product. Recrystallisation from ethyl acetate/ light petroleum gave the product as a white solid (1.29 g, 48%) identical in all respects to that produced by method 1.

Diethyl phthalimidomethylphosphonate (58).-[323]

N-bromomethylphthalimide (4.8 g, 20 mmol) and triethyl phosphite (3.32 g, 20 mmol) were stirred together in a flask fitted with a reflux condensor, and heated gently to initiate the exothermic reaction. When the exotherm had subsided the mixture was heated for 1 h., then dichloromethane (20 ml) was added. The resulting solution was then washed with water (3 x 10 ml), brine (10 ml), dried ($MgSO_4$) and evaporated under reduced pressure to give a yellow oil which crystallised on standing. Recrystallisation from dichloromethane/ light petroleum gave the product as a white solid (4.69 g, 79%) of m.p. 68-70°C (lit.[323] 67°C). ν_{max} (Nujol) 1719 (CO), 1240 (P=O), 1045 (P-O) and 782 cm^{-1} (aromatic); δ_H (200 MHz, C^2HCl_3) 1.24 (6H, app. t., *J* Hz, $C\underline{H}_3CH_2$ x2), 4.04 (2H, d., *J* Hz, PO-CH_2-N), 4.15 (4H, m., $C\underline{H}_2CH_3$) and 7.79 (4H, m., aromatic).

Diphenyl (N-benzyloxycarbonyl) aminomethanephosphonate (59).-[335]

A mixture of benzyl carbamate (1.53 g, 10 mmol), paraformaldehyde (300 mg, 10 mmol), acetic anhydride (1.19 ml, 12.5 mmol) and acetic acid (5 ml) was stirred at 65 °C for 3 h. Triphenyl phosphite (3.1 g, 10 mmol) was added and the mixture stirred at 120 °C for 2 h. Acetic acid and unreacted acetic anhydride were then removed under reduced pressure. The residue was dissolved in hot methanol and stored at -20 °C overnight. The resulting white solid was filtered, washed with methanol and recrystallised from chloroform/ methanol to give the product as a white solid (1.58 g, 32%) of m.p. 113-114 °C (lit.[335] 114-116 °C). *m/z* (Found: $(M + H)^+$

398.1155. $C_{21}H_{21}NO_5P$ requires 398.1152); ν_{max} (Nujol) 3273 (NH), 1721 (urethane CO), 1558 (amide II), 1226 (P=O), 953 (P-OPh) and 766 cm^{-1} (aromatic); δ_H (200 MHz, C^2HCl_3) 3.98 (2H, d., *J* 6.2 Hz, CH_2PO), 5.13 (2H, s., $PhCH_2O$), 5.32 (1H, br., NH) and 7.08-7.41 (15H, m., aromatic); δ_P (121.49 MHz, C^2HCl_3) 15.24 (3.8%) and 15.82 (96.2%); δ_C (50.31 MHz, C^2HCl_3) 35.97 & 39.15 (d., J_{PC} 160.3 Hz, CH_2-PO), 67.91 ($PhCH_2O$), 120.9-130.3 (aromatic) and 163.1 (urethane CO); *m/z* (CI) 415 ([*M* + NH_4]$^+$, 1.3%), 290 ([$OCNCH_2PO(OPh)_2$ + H]$^+$, 100), 264 ([*M*+H-Cbz]$^+$, 6.6), 94 (PhOH, 16.3) and 91 ($PhCH_2^+$, 33.1).

Diphenyl 1-aminomethanephosphonate hydrobromide (60).-

Compound (**59**) (7.94 g, 20 mmol) was dissolved in 30% hydrogen bromide in acetic acid (8 ml). The solution was stirred at room temperature for 1 h. and the solvent and volatiles were then removed under reduced pressure. The oily residue was crystallised from methanol and diethyl ether. Recrystallisation from methanol/ diethyl ether gave the product as white, hygroscopic crystals of m.p. 147-149°C (lit[335] 147-149°C). δ_H (200 MHz, ^2H_2O) 4.43 (2H, d., J_{PH} 10.3 Hz, CH_2) and 7.52-7.90 (10H, m., aromatic).

Diphenyl 1-aminobenzylphosphonate hydrobromide (61).-

N-Benzyloxycarbonyl protected (**34**) (4.87 g, 10 mmol) was dissolved in 30% hydrogen bromide in acetic acid (8 ml) and the solution stirred at room temperature for 1 h. After this time the volatiles were removed under reduced pressure to give an oily residue that was crystallised from methanol/ diethyl ether. Recrystallisation from methanol/ diethyl ether gave the product as white crystals (4.15 g, 96%) of m.p. 163.5-164.5 °C (lit.[332] 164-166 °C). δ_H (200 MHz, $C^2H_3O^2H$) 3.20 (1H, m., 1 of $C\underline{H}_2Ph$), 3.47 (1H, m., 1 of $C\underline{H}_2Ph$), 4.36 (1H, d.d.d., J_{CH} 6.8 and 8.3 Hz J_{PH} 14.6 Hz, CHPO) and 6.92-7.46 (15H, m., aromatic); δ_C (50.31 MHz, $C^2H_3O^2H$) 35.81 (CH_2Ph), 51.66 (d., J_{PC} 151.2 Hz, CH-P), 121.91-131.45 (aromatic) and 135.2-136.1 (quat. aromatic).

Diphenyl 1-aminobenzylphosphonate (62).-

Hydrobromide (**61**) (4.33g, 10 mmol) was suspended in chloroform (40 ml) and 2 M sodium hydroxide (40 ml) added. The mixture was shaken until all the solid had dissolved, the layers were separated and the organic layer dried ($MgSO_4$). The solvent was removed under reduced pressure to give a clear oil that solidified on standing. Recrystallisation from dichloromethane/ light petroleum gave the product as a white solid (3.20 g, 91%) of m.p. 61.5-62 °C. *m/z* (Found: $(M + H)^+$ 354.1259. $C_{20}H_{21}NO_3P$ requires 354.1259); ν_{max} (Nujol) 3285 (NH), 1278 (P=O), 932 (P-O) and 761 cm^{-1} (aromatic); δ_H (200 MHz, C^2HCl_3) 1.59 (2H, br.s., NH_2), 2.91 (1H, m., 1 of CH_2Ph), 3.45 (1H, m., 1 of CH_2Ph), 3.65 (1H, app. t. of d., J_{CH} 3.4 Hz J_{PH} 10.7 Hz, C<u>H</u>-PO) and 7.30 (15H, m., aromatic); δ_P (121.49 MHz, C^2HCl_3) 21.61; δ_C (50.31 MHz, C^2HCl_3) 38.32 (CH_2Ph), 49.67 & 52.71 (d., CH-PO, J_{PC} 153.1 Hz), 121.09-130.31 (aromatic) and 137.70 (quat. aromatic); *m/z* (CI) 354 ($[M + H]^+$, 4.6%), 235 ($[H_2PO(OPh)_2]^+$, 47.5) and 120 ($[M - PO(OPh)_2]^+$, 100).

Dibenzoyl (2S)-tartaric anhydride (63).-

To a stirred solution of freshly recrystallised dibenzoyl (2S)-tartaric acid (7.06 g, 20 mmol) in dry THF (50 ml) at 0 °C was added dropwise trifluoroacetic anhydride (2 ml) over 5 min. The mixture was then stirred at room temperature for 30 min., with the formation of white crystals. Removal of the solvent under reduced pressure gave white fluffy crystals which were recrystallised from ethyl acetate/ light petroleum to give the product as white crystals (6.32 g, 94%) of m.p. 194-6 °C. ν_{max} (Nujol) 1824 & 1740 (anhydride CO), 1707 (ester CO) and 713 cm^{-1} (aromatic); δ_H (200 MHz, C^2HCl_3) 5.98 (2H, s., CH x2), 7.42-7.75 (6H, m., aromatic) and 8.02-8.17 (4H, m., aromatic); δ_C (50.31 MHz, C^2HCl_3) 72.33 (CH), 126.61 (*ortho* aromatic), 128.24 (*para* aromatic), 129.79 (*meta* aromatic), 134.05 (*ipso* aromatic) and 171.64 (CO).

Resolution (64).-[367]

Diphenyl phosphonate ester (**62**) (7.08 g, 20 mmol) was dissolved in dry dioxane

(80 ml) and dibenzoyl-(2S)-tartaric anhydride (**63**) added (6.8 g, 20 mmol). The mixture was stirred until all solid dissolved and then left at room temperature for 3 days. The solvent was then removed under reduced pressure, benzene added (20 ml) and the solution left at 0 °C overnight. The (S)-diastereomer at C-1 of the phosphophenylalanine residue crystallised out preferentially. Reduction of the volume of benzene caused crystallisation of a second crop. The required (R) diastereomer was obtained by removing the benzene under reduced pressure, adding diethyl ether (20 ml) to the oil and standing at 0 °C for one day. A second crop was obtained by reduction of the volume of the solution to 10 ml and storing at 0 °C overnight.

Diphenyl 1-amino-*n*-butyl phosphonate hydrobromide (65).-

This was produced in an identical manner to (**61**) using N-protected (**48**) (4.59 g, 10 mmol). Recrystallisation from methanol/ diethyl ether gave the product as white crystals (3.85 g, 95%) of m.p. 172-173 °C. δ_H (200 MHz, $C^2H_3O^2H$) 1.16 (3H, t., *J* 7.1 Hz, CH_3), 1.31-1.79 (4H, m., CH_2CH_2), 1.88-2.37 (2H, m., $C\underline{H}_2CHP$), 4.14 (1H, d.d., J_{CH} 4.2 Hz, J_{PH} 13.3 Hz, CHPO) and 7.17-7.52 (10H, m., aromatic); δ_P (121.49 MHz, $C^2H_3O^2H$) 18.67; δ_C (50.31 MHz, $C^2H_3O^2H$) 14.32 (CH_3), 23.63 ($CH_3\underline{C}H_2$), 29.14 (d., J_{PC} 8.54 Hz, $CH_3CH_2\underline{C}H_2$), 29.59 (d., J_{PC} 2.64 Hz, $\underline{C}H_2CH$), 47.31 & 50.46 (d., J_{PC} 158.5 Hz, CHPO), 121.93 & 122.03 (*ortho* aromatic), 127.62 (*para* aromatic) and 131.51 (*meta* aromatic).

Diphenyl 1-amino-2-methylpropylphosphonate hydrobromide (66).-

This was prepared in an identical manner to compound (**61**) using the N-protected compound (**49**) (4.59 g, 10 mmol). Recrystallisation from methanol/ diethyl ether gave the product as white crystals (3.78 g, 92%) of m.p. 174-176 °C. δ_H (200 MHz, $C^2H_3O^2H$) 1.12 (6H, d., *J* 5.9 Hz, $(CH_3)_2$), 1.83-2.19 (3H, m., $(CH_3)_2C\underline{H}C\underline{H}_2$), 4.24 (1H, m., CHPO) and 7.08-7.52 (10H, m., aromatic); δ_P (121.49 MHz, $C^2H_3O^2H$) 18.13; δ_C (50.31 MHz, $C^2H_3O^2H$) 22.44† (1 of $(\underline{C}H_3)_2$)CH), 23.23† (1 of $(\underline{C}H_3)_2$)CH), 25.94 (d., J_{PC} 9.6 Hz, $(CH_3)_2\underline{C}H$), 38.57 and 35.59 (conformers of $\underline{C}H_2CHP$), 45.58

& 48.74 (d., J_{PC} 158.9 Hz, $\underline{C}$HPO) 122.02 & 123.41 (*ortho* aromatic), 127.65 (*para* aromatic), 131.56 (*meta* aromatic) and 152.71 (*ipso* aromatic).

Diphenyl 1-amino-*iso*propyl phosphonate hydrobromide (67).-

This was prepared in an identical manner to (**61**) using the N-protected compound (**50**) (4.33 g, 10 mmol). Recrystallisation from methanol/ diethyl ether gave the product as white crystals (3.54 g, 86%) of m.p. 173-5 °C. δ_H (200 MHz, $C^2H_3O^2H$) 1.16† (6H, d.d.'s, J_{CH} 6.9 Hz and J_{PH} 1.2 Hz, $(C\underline{H}_3)_2CH$), 1.56 (2H, s., NH_2), 2.37 (1H, d.h., J 4.2 and 6.9 Hz, $(CH_3)_2C\underline{H}$), 3.23 (1H, d.d., J_{CH} 4.2 and J_{PH} 13.3 Hz, CHPO) and 7.08-7.47 (10H, m., aromatic); δ_P (121.49 MHz, $C^2H_3O^2H$) 19.34; δ_C (50.31 MHz, $C^2H_3O^2H$) 17.88† ($(CH_3)_2$), 21.18† ($(CH_3)_2$), 29.75 (d., J_{PC} 6.1 Hz, $(CH_3)_2\underline{C}H$), 53.71 and 56.57 (d., J_{PC} 143.5 Hz, CHPO), 121.04 (*ortho* aromatic), 125.53 (*para* aromatic), 130.22 (*meta* aromatic) and 153.21 (*ipso* aromatic).

Diphenyl 1-amino-*n*-butyl phosphonate (68).-

This was prepared in an identical manner to (**62**) using the *n*-butyl hydrobromide (**65**) (3 g, 7.5 mmol) to give the product as a clear oil (1.96 g, 82%). *m/z* (Found: $(M+H)^+$ 320.3514. $C_{17}H_{23}NO_3P$ requires 320.3515); ν_{max} (Nujol) 3304 (NH), 1266 (P=O), 938 (P-O) and 765 cm^{-1} (aromatic); δ_H (200 MHz, C^2HCl_3) 0.94 (3H, t., J 7.1 Hz, $C\underline{H}_3CH_2$),1.22-1.72 (5H, m., $CH_3C\underline{H}_2C\underline{H}_2$- and 1 of $C\underline{H}_2CHP$), 1.65 (2H, s., NH_2), 2.04 (1H, m., 1 of $C\underline{H}_2CHP$), 3.32 (1H, app. d.t., J_{CH} 3.95 and J_{PH} 9.4 Hz, CHPO) and 7.08-7.47 (10H, m., aromatic); δ_P (121.49 MHz, C^2HCl_3) 22.01; δ_C (50.31 MHz, C^2HCl_3) 14.31 ($\underline{C}H_3CH_2$), 23.02 ($CH_3\underline{C}H_2$), 29.75 (d., J 6.8 Hz, $CH_3CH_2\underline{C}H_2$), 31.7 (d., J 2.9 Hz, $\underline{C}H_2CHP$), 48.23 and 51.21 (d., J_{PC} 152.5 Hz, CHPO), 121.06 and 121.0 (*ortho* aromatic), 125.53 (*para* aromatic), 130.22 (*meta* aromatic) and 153.83 (*ipso* aromatic); *m/z* (CI) 320 ($[M+H]^+$, 24.8%) and 87 $[M-PO(OPh)_2]^+$, 100).

Diphenyl 1-amino-2-methylpropylphosphonate (69).-

This was synthesised in an identical manner to (**62**) using hydrobromide (**66**) (3.8 g, 9.5 mmol) giving the product as a clear oil (2.75 g, 91%). *m/z* (Found: $(M + H)^+$ 320.3514. $C_{17}H_{23}NO_3P$ requires 320.3515); ν_{max} (Nujol) 3295 (NH), 1286 (P=O), 927 (P-O) and 760 cm^{-1} (aromatic); δ_H (200 MHz, C^2HCl_3) 0.92 (3H, d., *J* 6.5 Hz, 1 of $(CH_3)_2$), 1.01 (3H, d., *J* 6.5 Hz, 1 of $(C\underline{H}_3)_2$), 1.55-1.87 (2H, m., $(CH_3)_2C\underline{H}$ and 1 of CH_2), 1.62 (2H, s., NH_2), 2.01 (1H, m., 1 of CH_2), 3.41 (1H, m., CHPO) and 7.04-7.47 (10H, m., aromatic); δ_P (121.49 MHz, C^2HCl_3) 21.92; δ_C (50.31 MHz, C^2HCl_3) 21.52† ($(\underline{C}H_3)_2CH$), 24.01† ($(\underline{C}H_3)_2CH$), 24.59 & 24.87† ($(CH_3)_2\underline{C}H$), 40.29 (d., J_{PC} 16.8 Hz, $\underline{C}H_2CHP$), 46.40 & 49.30 (d., J_{PC} 146.1 Hz, CHPO), 120.96 (*ortho* aromatic), 125.58 (*para* aromatic), 130.23 (*meta* aromatic) and 151.20 (*ipso* aromatic); *m/z* (CI) 320 ($[M+H]^+$, 20.4%) and 87 $[M - PO(OPh)_2]^+$, 100).

Diphenyl 1-amino-*iso*propyl phosphonate (70).-

This was prepared in an identical manner to compound (**63**) using the hydrobromide (**68**) (2.9 g, 7.5 mmol) giving the product as a clear oil (1.81 g, 79%). *m/z* (Found: $(M + H)^+$ 306.3245. $C_{16}H_{21}NO_3P$ requires 306.3244); ν_{max} (Nujol) 3297 (NH), 1271 (P=O), 922 (P-O) and 757 cm^{-1} (aromatic); δ_H (200 MHz, C^2HCl_3) 1.15† (6H, d., *J* 6.9 Hz, $(C\underline{H}_3)_2CH$), 1.17† (1H, d.d., *J* 6.9 and 1.2 Hz, $(C\underline{H}_3)_2CH$), 1.56 (2H, s., NH_2), 2.37 (1H, d.h., *J* 4.1 and *J* 6.9 Hz, $C\underline{H}(CH_3)_2$), 3.23 (1H, d.d., J_{CH} 4.1 and J_{PH} 13.3 Hz, CHPO) and 7.08-7.47 (10H, m., aromatic); δ_P (121.49 MHz, C^2HCl_3) 21.71; δ_C (50.31 MHz, C^2HCl_3) 17.89† ($(\underline{C}H_3)_2CH$), 21.19† ($(\underline{C}H_3)_2CH$), 29.76 (d., J_{PC} 11.6 Hz, $(CH_3)_2\underline{C}H$), 53.72 & 56.57 (d., J_{PC} 143.5 Hz, CHPO), 120.97 & 121.06 (*ortho* aromatic), 125.53 (*para* aromatic), 130.22 (*meta* aromatic) and 152.64 (*ipso* aromatic); *m/z* (CI) 306 ($[M + H]^+$, 19.4%) and 73 $[M - PO(OPh)_2]^+$, 100).

(2S, 3R) N-Benzyloxycarbonyl-threonine (71).-

(2S, 3R)-Threonine (0.6 g, 5 mmol) was dissolved in water (50 ml) containing $NaHCO_3$ (1.05 g, 12.5 mmol) with vigorous stirring. Benzyloxychloroformate (0.8 ml,

5.5 mmol) was added in five portions over 15 min. The mixture was stirred for 3 h. at room temperature, washed with diethyl ether (2 x 25 ml), acidified with concentrated HCl to pH 1 and extracted with diethyl ether (3 x 25 ml). The ethereal extracts were washed with brine (2 x 25 ml), dried (Na_2SO_4) and evaporated under reduced pressure. The resulting oil was crystallised from ethyl acetate/ light petroleum to give white crystals (915 mg, 72%) of m.p. 101-103 °C (lit.[454] 101-3 °C). *m/z* (Found: $(M+H)^+$ 254.1025. $C_{12}H_{15}NO_5$ requires 254.1028); $[\alpha]_D^{20}$ -4.1° (c 4 in AcOH) (lit.[454] $[\alpha]_D^{20}$ -4.7° (c 4 in AcOH)); ν_{max} (Nujol) 3402 (NH), 1724-1710 (CO) and 1525 cm^{-1} (aromatic); δ_H (90 MHz, d_6-DMSO) 1.09 (3H, d., *J* 6.9 Hz, CH_3), 4.3 (2H, m., C-2 & C-3 H), 4.95 (2H, s., PhC$\underline{H}_2$), 6.07 (1H, d., *J* 9.1 Hz, NH), 7.07 (1H, br.s., OH), 7.18 (5H, s., C_6H_5); δ_C (50.31 MHz, d_6-DMSO) 20.63 (CH_3CH), 60.24 (CH_3O), 65.81 (CHNH), 66.73 (CHOH), 127.93, 128.09 & 128.63 (aromatic), 137.24 (quat. aromatic), 156.69 (urethane CO) and 172.64 (carboxyl CO); *m/z* (EI) 253 (M^+, 2.1%), 209 ($[M-CO_2]^+$, 4.5), 108 ($C_7H_7OH^+$, 92.1), 91 ($C_6H_5CH_2^+$, 38.4) and 79 ($C_6H_7^+$, 100).

Methyl (2S, 3R) N- benzyloxycarbonyl threoninate (72).-

To an ice-cold solution of (**71**) (13.5 g, 53 mmol) in diethyl ether (70 ml) was added excess ethereal diazomethane (6.7 g, 0.16 mol) with stirring and the reaction allowed to warm to room temperature over 1 h. Unreacted diazomethane was removed in a stream of nitrogen and the solvent removed under reduced pressure to give a yellowish solid which was recrystallised from ethyl acetate/ light petroleum to give white crystals (10.6 g, 78%) of m.p. 91-92.5 °C (lit.[304] 88-89 °C). *m/z* (Found: $(M+H)^+$ 268.1185. $C_{13}H_{18}NO_5$ requires 268.1184); $[\alpha]_D^{20}$ -14.5° (c 1 in $CHCl_3$) (lit.[304] $[\alpha]_D^{20}$ -13.7° (c 1 in $CHCl_3$)); ν_{max} (Nujol) 3446 (NH), 1716 (ester CO), 1688 (amide CO), 1545 (aromatic) and 1300-1200 cm^{-1} (C-O); δ_H (200 MHz, C^2HCl_3) 1.21 (3H, d., *J* 6.5 Hz, CH_3), 3.74 (3H, s., OCH_3), 4.36 (1H, d.q., *J* 6.5 Hz, *J* 3.6 Hz, CHOH), 4.58 (1H, d.d., J_{CH} 3.6 Hz, J_{NH} 8.5 Hz, C$\underline{H}$NH), 5.14 (2H, s., PhC$\underline{H}_2$O), 5.71 (1H, d., *J* 8.5 Hz, NH) and 7.35 (5H, s., aromatic); δ_C (50.31 MHz, C^2HCl_3) 19.76 (CH_3), 52.42 ($CO_2\underline{C}H_3$), 59.59 (CH), 67.09 (CH_2Ph), 67.81 (CHNH), 127.96, 128.13 & 128.53 (aromatic), 156.41 (urethane CO) and 171.67 (ester CO); *m/z* (EI) 267 (M^+,

0.5%), 223 ([*M*-CO_2]$^+$, 4.6), 108 ($C_7H_7OH^+$, 58.9) and 91 ($C_6H_5CH_2^+$, 100).

Methyl (2R, 3S)-(N- benzyloxycarbonyl)-2-amino-3-bromobutyrate (73).-

Carbon tetrabromide (12.3 g, 37 mmol) and triphenylphosphine (9.71 g, 37 mmol) (both dried over phosphorous pentoxide for 12 h.) were dissolved in dry benzene (230 ml). The methyl ester (**72**) (6.9 g, 26 mmol) was added in several portions over 30 min. at room temperature. The reaction was stirred for a further 90 min., after which time the triphenylphosphine oxide was filtered off and the pad washed with dry benzene (75 ml). The combined filtrates were concentrated under reduced pressure and the resulting oil was purified by flash column chromatography (3:1 ethyl acetate/ chloroform) and then recrystallised twice from ethyl acetate/ light petroleum to give the product as white crystals (2.88 g, 49%) of m.p 73-76 °C (lit.[304] 73-74 °C). *m/z* (Found: (*M* + H)$^+$ 331.0341. $C_{13}H_{17}NO_4{}^{79}Br$ requires 331.0341); $[\alpha]_D^{20}$ + 28.9 ° (c 1 in $CHCl_3$) (lit.[304] $[\alpha]_D^{20}$ +31.5° (c 0.47 in $CHCl_3$)); ν_{max} (Nujol) 3301 (NH), 1740 (ester CO), 1684 (amide CO), 1533 (aromatic) and 1300-1100 cm^{-1} (C-O); δ_H (200 MHz, C^2HCl_3) 1.71 (3H, d., *J* 7.4 Hz, CH_3), 3.79 (3H, s., OCH_3), 4.35 (1H, d.q., *J* 7.4 & 3.6 Hz, C-2 H), 4.61 (1H, d.d., J_{CH} 3.6, J_{NH} 8.5 Hz, C-3 H), 5.13 (2H, s., $PhCH_2O$), 5.72 (1H, d., *J* 8.5 Hz, NH), 7.38 (5H, s., aromatic); δ_C (50.31 MHz, C^2HCl_3) 23.23 ($\underline{C}H_3CH$), 49.72 (CHBr), 53.2 (CH_3O), 60.23 (CHNH), 67.85 (CH_2Ph), 128.65-129.07 (aromatic), 136.44 (*ipso* aromatic), 155.9 (urethane CO) and 169.53 (ester CO); *m/z* (EI) 331 & 329 (*M*$^+$, Br isotopes, 0.6 & 0.6%), 272 ([*M*-CO_2Me]$^+$, 4.6), 108 ($C_7H_7OH^+$, 42.8) and 91 ($C_6H_5CH_2^+$, 100).

(2R, 3S)-2-amino-3-bromobutyric acid hydrochloride (74).-

The fully protected bromobutyrine (**73**) (1.4 g, 4.23 mmol) was refluxed in a mixture of 12 M HCl/ glacial acetic acid (35 ml, 6:4 v/v) for 2.5 h. The solvent was removed under reduced pressure to give a pale yellow solid. This solid was recrystallised from methanol/ diethyl ether to give white crystals (425 mg, 48%) of m.p 187-188 °C (dec.) (lit.[304] 188 °C). $[\alpha]_D^{20}$ +13.1° (c 0.5 in H_2O) (lit.[304] $[\alpha]_D^{20}$ +12.7° (c 0.44 in H_2O)); ν_{max} (Nujol) 3051 (NH_3^+), 1728 (CO) and 1280-1100 cm^{-1} (C-O); δ_H (270 MHz,

^2H_2O) 1.78 (3H, d., *J* 7 Hz, $CHC\underline{H}_3$), 4.20 (1H, d., *J* 2.9 Hz, C-2 H) and 4.68 (1H, q., *J* 2.9 Hz, $C\underline{H}CH_3$); δ_C (75.74 MHz, ^2H_2O) 19.25 ($CH\underline{C}H_3$), 43.72 (CHNH), 57.88 ($\underline{C}HCH_3$) and 167.34 (CO); *m/z* (FAB) 184, 182 ([*M*+ H-Cl]$^+$, bromine isotopes, 100%) and 102 ([*M*-Br]$^+$, 10.6).

(2S, 3S)-3-Methylaspartic acid (75).-

Mesaconic acid (13 g, 0.1 mol) was suspended in water (100 ml) and the pH adjusted to 9 with concentrated NH_3. The solution was then concentrated *in vacuo*. The resulting diammonium mesaconate was redissolved in water (125 ml) and magnesium chloride hexahydrate (250 mg, 62.5 mmol) and potassium chloride (45 mg, 28.5 mmol) were added, along with the enzyme 3-methylaspartase (150 units). The reaction was incubated at 30 °C and aliquots (20 µl) were taken at intervals and the absorbance at 240 nm measured until no further decrease was observed (46 h.). The protein was denatured by boiling the solution for 2 min. and removed by filtration. The filtrate was evaporated to dryness, the residue dissolved in water (30 ml) and the pH was adjusted to approximately 3.5 with 12 M HCl. The product spontaneously precipitated as a yellow solid and was recrystallised twice from water/ ethanol to give white crystals (8.4 g, 55.7%) of m.p 270-273 °C (lit.[455] 276-278 °C). $[\alpha]_D^{20}$ -9.2° (c 0.6 in H_2O) (lit.[455] $[\alpha]_D^{20}$ -10° (c 0.42 in H_2O)); ν_{max} (Nujol) 2924 (NH), 1866 & 1681 (CO) and 1611 cm^{-1}; δ_H (90 MHz, ^2H_2O) 1.13 (3H, d., *J* 8 Hz, CH_3), 3.08 (1H, m., C-3 H) and 4.00 (1H, d., *J* 3 Hz, C-2 H); δ_C (50.31 MHz, ^2H_2O, pD 1) 12.86 (CH_3), 38.99 (CH), 54.16 (CHNH) and 167.21 & 174.9 (CO); *m/z* (FAB) 148 ([*M*+H]$^+$, 5.9%).

α-Methyl N-trifluoroacetyl-(2S, 3S)-3-methylaspartate (76).-

Trifluoroacetic anhydride (48.5 g, 231 mmol) was added to a stirred suspension of (**75**) (4 g, 27.2 mmol) under nitrogen in dry THF (125 ml) at 0 °C over 30 min. The reaction was allowed to warm to room temperature and was left stirring for 2 h., when dissolution was complete. The solvent was removed *in vacuo* and the resulting anhydride (6.1 g, 99%) thoroughly dried (phosphorous pentoxide). The anhydride

was treated with dry methanol (50 ml) at 0 °C and the solution was allowed to warm to room temperature over 2 h. The solvent was removed under reduced pressure to give an off-white solid (6.75 g, 96%) (a mixture of the α & β methyl esters); δ_H (90 MHz, C^2HCl_3) 1.36 (3H, d., *J* 7.3 Hz, CH_3), 3.07 (1H, m., C-3 H), 3.75 & 3.83 (β & α methyl esters respectively), 4.89 (1H, m., C-2 H) & 7.26 (1H, br.s., NH).

α-Methyl N-trifluoroacetyl-(2S, 3S)-3-methylaspartyl-β-acid chloride (77).-

The mixture of monoesters (**76**) from the above reaction (5 g, 19.4 mmol) were heated to reflux in thionyl chloride (60 ml) for 1 h. The solvent was removed under reduced pressure to give a pale yellow solid, which would not recrystallise from diethyl ether/ light petroleum. Consequently the crude product (4.34 g, 81%) was taken to the next step as an oil and the separation of the two diastereomeric ester/ acid chlorides was not achieved. δ_H (90 MHz, C^2HCl_3) 1.31 (3H, d., *J* 7Hz, $C\underline{H}_3CH$), 3.53 (1H, m., C-3 H), 3.67 & 3.83 (β & α OMe respectively) 4.89 (1H, m., C-2 H) and 7.31 (1H, br. s., NH).

α-Methyl N-trifluoroacetyl-(2S, 3S)-3-methylaspartyl-*m*-chlorobenzoyl peroxide (78).-

The N-trifluoroacetyl-(2S, 3S)-3-methylaspartyl-α & β chlorides (**77**) (1 g of the crude oil) was dissolved in dry diethyl ether (30 ml) and 80-90% *m*-chloroperbenzoic acid (734 mg, 3.62 mmol) added and the mixture stirred under nitrogen in an ice/salt bath. Pyridine (246 mg, 3.6 mmol) in dry diethyl ether (2 ml) was added dropwise to the mixture and stirring continued overnight. The solution was filtered and washed with water (2 x 15 ml) and 1 M aqueous sodium carbonate (2 x 15 ml), dried (Na_2SO_4) and the solvent removed under reduced pressure to give a white solid. A fractional crystallisation was then attempted, but this produced a white solid which was a mixture of the two diastereomers. Consequently this synthesis of L-*allo*-threonine was abandoned.

Methyl (2S, 3S)-N-benzyloxycarbonyl threoninate (79).-

This was prepared in an identical manner to the (2S, 3R) diastereomer, but starting from (2S, 3S)-threonine (100 mg, 0.804 mmol). The crude (2S,3S)-N-benzyloxycarbonyl threonine was identified only by ^{1}H n.m.r. δ_H (90 MHz, C^2HCl_3) 1.21 (3H, d., *J* 7 Hz, CH_3), 4.22 (2H, m, C-2 & C-3 H), 5.05 (2H, s., $PhCH_2O$), 5.71 (1H, s., OH), 6.31 (1H, d., *J* 9 Hz, NH) and 7.27 (5H, s., aromatic). The N-protected (2S, 3S)-threonine was dissolved in diethyl ether and treated with excess diazomethane as before, (**73**). Recrystallisation from diethyl ether/ light petroleum yielded white crystals of the methyl ester (103 mg, 46% over two steps) of m.p 57-58.5 °C (lit.[304] 58-59 °C). $[\alpha]_D^{20}$ +12.1° (c 0.4 in $CHCl_3$) (lit.[304] $[\alpha]_D^{20}$ +13.7 ° (c 0.5 in $CHCl_3$)); ν_{max} ($CHCl_3$) 3430 (NH), 3030-2970 (CH, CH_2, CH_3), 1755-1700 (ester & amide CO), 1510 (aromatic) and 1300-1100 cm^{-1} (C-O); δ_H (90 MHz, C^2HCl_3) 1.21 (3H, d., *J* 7Hz, 3-CH_3), 2.50 (1H, s., OH), 3.76 (3H, s., OCH_3), 4.12 (1H, m., C-3 H), 4.39 (1H, m., C-2 H), 5.11 (2H, s., $PhCH_2O$), 5.68 (1H, d., *J* 9.1 Hz, NH) and 7.34 (5H, s., aromatic); δ_C (89.71 MHz, C^2HCl_3) 19.11 ($CH\underline{C}H_3$), 52.58 (OCH_3), 59.88 (CHOH), 67.32 (OCH_2), 69.23 (CHNH), 128.24, 128.43 & 128.75 (aromatic), 156.68 (urethane CO) and 171.03 (ester CO); *m/z* (EI) 267 (M^+, 0.2%), 223 ([*M*-CO_2]$^+$, 4.2), 108 ($C_7H_7OH^+$, 58.9) & 91 ($C_6H_5CH_2^+$, 100).

Methyl (2R, 3R)-(N-benzyloxycarbonyl)-2-amino-3-bromobutyrate (80).-

This was prepared in an identical manner to the (2S, 3R) epimer, from methyl ester (**79**) (100 mg, 0.36 mmol) except that further purification of the oil produced by the ethyl acetate: chloroform column was necessary. Thus a second flash column was performed, using chloroform as the eluant to give a clear oil which could not be crystallised (56 mg, 45%). $[\alpha]_D^{20}$ + 2.6° (c 0.8 in $CHCl_3$) (lit.[304] $[\alpha]_D^{20}$ + 3.1 °) (c 1.76 in $CHCl_3$); ν_{max} (Nujol) 3305 (NH), 1737 (ester CO), 1680 (amide CO), 1533 (aromatic) and 1300-1100 cm^{-1} (C-O); δ_H (90 MHz, C^2HCl_3) 1.65 (3H, d., *J* 7 Hz, 3-CH_3), 3.69 (3H, s., OCH_3), 4.67 (2H, m., C-2 & C-3 H), 5.05 (2H, s., $PhCH_2O$), 5.6 (1H, d., *J* 9 Hz, NH) and 7.37 (5H, s., aromatic); *m/z* (EI) 331 & 329 (M^+, Br isotopes, 0.6 & 0.6%), 272 ([*M*-CO_2Me]$^+$, 5.9 %), 108 ($C_7H_7OH^+$, 84.3) & 91 ($C_6H_5CH_2^+$, 100).

(2R, 3R)-2-amino-3-bromobutyric acid hydrochloride (81).-

This was synthesised exactly as for its 3R epimer, from 55 mg (0.166 mmol) of (**80**) giving white crystals (20 mg, 55%); δ_H (90 MHz, ^2H_2O) 1.64 (3H, d., *J* 7 Hz, CH_3), 4.18 (1H, d., *J* 2.9 Hz, C-2 H) and 4.64 (1H, m., C-3 H);

Methyl (2S, 3R)-N-*tert*-butoxycarbonyl-threoninate (82).-

This was prepared as for the benzyloxycarbonyl analogue (**72**), using N-*tert*-butoxycarbonyl-(2S)-threonine (4 g, 18.3 mmol). After evaporation of the solvent the oily residue was purified by flash chromatography (ethyl acetate) to give the product as a clear oil (4.11 g, 96%). $[\alpha]_D^{20}$ + 17.3° (c 1 in $CHCl_3$); δ_H (200 MHz, C^2HCl_3) 1.22 (3H, d., *J* 6.3 Hz, $C\underline{H}_3CH$), 1.44 (9H, s., $(CH_3)_3C$), 2.23 (1H, s., OH), 3.74 (3H, s., OCH_3), 4.28 (2H, m., C-2 and C-3 CH) and 5.89 (1H, br.s. NH); δ_C (74.76 MHz, C^2HCl_3) 20.55 ($\underline{C}H_3CH$), 28.95 (($\underline{C}H_3)_3C$), 53.17 (OCH_3), 61.10 (CH-NH), 68.76 (CHOH), 80.79 (($CH_3)_3\underline{C}$), 157.40 (urethane CO) and 172.69 (ester CO).

Methyl (2R, 3R) N-*tert*-butoxycarbonyl-2-amino-3-bromobutyrate (83).-

This was prepared as for the benzyloxycarbonyl analogue, using ester (**82**) (4.01 g, 17 mmol). After filtration and evaporation of the solvent the oily residue was purified by flash chromatography (ethyl acetate) to give the product as a clear oil (2.11 g, 42%). δ_H (200 MHz, C^2HCl_3) 1.45 (9H, s., $((CH_3)_3C$), 1.79 (3H, d., *J* 6.56 Hz, $C\underline{H}_3CH$), 3.79 (3H, s., OCH_3), 4.32 (1H, q.d., *J* 3.5 & 6.56 Hz, CHBr), 4.54 (1H, d.d., J_{NH} 8.5 Hz J_{CH} 3.5 Hz, $C\underline{H}NH$) and 5.41 (1H, br., NH).

Active-site model reactions (84).-

Succinic anhydride (114 mg, 1 mmol) was dissolved in dry THF (10 ml) and *p*-nitrophenylhydrazine (153 mg, 1 mmol) added in portions over 5 min. After 100 h. the solvent was evaporated, the residue taken up in chloroform (20 ml), filtered and the

filtrate washed with saturated $NaHCO_3$ (4 x 10 ml), dried and the solvent removed under reduced pressure to give a dark red solid. ν_{max} ($CHCl_3$) 1736 (CO), 1599 (aromatic), 1525 (NO_2) and 736 cm^{-1} (aromatic, *p*-substituted). δ_H (90 MHz, C^2HCl_3) 2.05 (4H, m., CH_2 x2), 7.08 (2H, d., *J* 7.5 Hz, aromatic), 7.5 (1H, br.s, NH) and 8.15 (2H, d., *J* 7.5 Hz, aromatic);

Phthalic anhydride (148 mg, 1 mmol) was dissolved in dry THF (40 ml) and *p*-nitrophenylhydrazine (153 mg, 1 mmol) was added in portions over 5 min. Almost instantly a cream-coloured solid began to precipitate, and precipitation appeared to be complete after 30 min. The mixture was filtered, and the solid washed with dry THF (10 ml) and dried, to give the product as a cream solid (1.84 g, 65%) of m.p. 241-243 °C; ν_{max} (Nujol) 1672 (amide CO), 1648 (acid CO), 1592 (aromatic) and 800-750 cm^{-1} (aromatic), δ_H (200 MHz, C^2HCl_3) 7.1 (4H, m., aromatic), 7.5 (1H, br.s, NH) and 8.05 (4H, m., aromatic); *m/z* (EI) 283 (M^+, 100%), 104 ($[C_6H_4\text{-}CO]^+$, 71.1) and 76 ($C_6H_4^+$, 60.8)

Resolution using dibenzoyl-(D)-tartaric acid (85).-[354b]

Diphenyl 1-aminobenzyl phosphonate (**62**) (1.06 g, 3 mmol) was dissolved in ethanol/ methanol (4:1, 5 ml) and dibenzoyl D-(+) tartaric acid monohydrate (1.13 g, 3 mmol) in ethanol (15 ml) added. After 3 h. stirring at room temperature the suspension was filtered and washed with ethanol to give a white solid that was recrystallised twice from ethanol to give a white solid (1.02 g, 73 %) of m.p. 175-177 °C. The aminophosphonic acid was liberated from this salt by dissolving in 1 M sodium hydroxide (10 ml), saturating the aqueous layer with sodium chloride and extracting with chloroform (3 x 15 ml). The organic extracts were dried ($MgSO_4$) and the solvent removed under reduced pressure to give a white solid in all respects identical to compound (**62**) except that it has an optical rotation of - 45° (c 1 in MeOH).

Lithium tri-*tert*-butoxy aluminium hydride (86).-[341]

Lithium aluminium hydride (4.56 g, 120 mmol) was suspended in 120 ml dry diethyl ether and stirred for 30 min. at room temperature, after which time the mixture was filtered under an inert atmosphere through Celite. To the stirred filtrate was added dry *tert*-butyl alcohol (26.64 g, 360 mmol) in dry diethyl ether (150 ml) dropwise under nitrogen over 30 min. The resuting white solid was dried under reduced pressure and used directly.

Triphenylmethylamine (87).-[369]

Triphenylmethyl chloride (10 g, 36 mmol) was suspended in concentrated aqueous ammonia (50 ml) and stirred at room temperature for 2 days. The resulting suspension was filtered and the solid recrystallised from toluene/ light petroleum to give the product as a cream crystals (5.95 g, 64%) of m.p. 101-3 °C (lit.[369] 103-104 °C). δ_H (200 MHz, C^2HCl_3) 2.42 (2H, br., NH_2) and 7.32 (15H, s., aromatic); δ_C (50.31 MHz, C^2HCl_3) 66.77 ($Ph_3\underline{C}$-NH_2), 127.12 (*para* aromatic), 128.48 (*ortho* aromatic), 128.71 (*meta* aromatic) and 149.20 (*ipso* aromatic); *m/z* (EI) 231 (Ph_3C^+, 100).

Triphenylmethanimine (88).- Method 1

Amine (**87**) (2.59 g, 10 mmol) and paraformaldehyde (600 mg, 20 mmol) were dissolved in toluene (25 ml) and stirred at room temperature for 24 h. After this time magnesium sulfate was added and the mixture stirred for 20 min. then filtered. The filtrate was evaporated under reduced pressure to give a white solid that was recrystallised from toluene/ light petroleum (1.92 g, 71%). The compound was identical in all respects to that obtained from method 2.

Method 2[369]

Amine (**87**) (2.59 g, 10 mmol) was dissolved in toluene (20 ml) and aqueous formaldehyde (1.7 g, 20 mmol) added. The mixture was vigorously stirred for 24 h. at room temperature and then sodium sulfate added and the suspension stirred for a further 20 min. The mixture was filtered and the filtrate evaporated under reduced

pressure to give a cream crystalline solid. Recrystallisation from toluene/ light petroleum gave the product as white crystals (2.15 g, 79%) of m.p. 127-9 °C (lit.[369] 134-6 °C). (Found: C, 88.81; H, 6.32; N, 5.16. Calc. for $C_{20}H_{11}N$: C, 88.52; H, 5.81; N, 5.16); δ_H (200 MHz, C^2HCl_3) 7.05 (1H, d., *J* 15.9 Hz, 1 of CH_2=N), 7.19-7.29 (15H, s., aromatic) and 7.80 (1H, d., *J* 15.9 Hz, 1 of CH_2=N); δ_C (50.31 MHz, C^2HCl_3) 80.60 ($Ph_3\underline{C}$), 127.81 (*para* aromatic), 128.50 (*ortho* aromatic), 130.37 (*meta* aromatic), 145.29 (*ipso* aromatic) and 153.79 (C=N); *m/z* (CI) 231 (Ph_3C^+, 100%).

Dimethyl 1-(triphenylmethylamino)methylphosphonate (89).-

N-Triphenylmethanimine (**88**) (2.71 g, 10 mmol) was dissolved in toluene (25 ml), dimethyl phosphite (5.5 g, 50 mmol) added and the mixture heated to 100 °C. After 5 min. a white solid was deposited and after a further 25 min. the reaction was cooled, filtered and the solid recrystallised from chloroform/ methanol to give the product as white solid (2.59 g, 66%) of m.p. 208-9 °C (lit.[369] 210-11 °C). *m/z* (Found: $(M + H)^+$ 381.1494. $C_{22}H_{24}NO_3P$ requires 381.14937); ν_{max} (Nujol) 3326 (NH), 1238 (P=O), 1014 (P-OC) and 761 cm^{-1} (aromatic); δ_H (200 MHz, C^2HCl_3) 2.04 (1H, br., NH), 2.57 & 2.54 (1H, d., J_{PH} 13.83 Hz, 1 of CH_2PO), 3.87 (6H, d., J_{PH} 10.65 Hz, $PO(OCH_3)_2$) and 7.15-7.61 (15H, s., Ph_3C); δ_P (121.49 MHz, C^2HCl_3) 30.12; δ_C (50.31 MHz, C^2HCl_3) 38.93 & 41.05 (d., J_{PC} 160.7 Hz, CH_2PO), 53.83 (d., J_{PC} 6.82 Hz, $PO(OCH_3)_2$), 127.48 (*para* aromatic), 128.90 (*ortho* aromatic), 129.37 (*meta* aromatic) and 145.57 (quat. aromatic); *m/z* (CI) 381 (M^+, 4%), 304 ([*M*-Ph]$^+$, 59.8), 258 (Ph_3CNH^+, 3), 243 (Ph_3C^+, 100), 110 ($H_2NCH_2PO(OCH_3)_2^+$, 2) and 91 ($PhCH_2^+$, 11.3).

2-Hydroxy-N-benzyloxycarbonyl glycine (90).-[349]

Freshly recrystallised glyoxylic acid monohydrate (19.6 g, 0.21 mol) was dissolved in dry diethyl ether (200 ml) and benzyl carbamate (28.9 g, 0.19 mol) added. The mixture was stirred overnight at room temperature and the resulting solid filtered and washed with diethyl ether to give a white solid. Recrystallisation from ethyl acetate/ light petroleum gave the product as a white solid (25.6 g, 63.5%) of m.p. 193-195 °C

(lit.[349] 196-8 °C). ν_{max} (Nujol) 3324 (NH), 1731 (acid CO), 1696 (urethane CO) and 699 cm^{-1} (aromatic); δ_H (200 MHz, d_6-DMSO) 5.06 (2H, s., $PhCH_2O$), 5.24 (1H, d., *J* 6.9 Hz, CH), 7.31 (5H, s., aromatic) and 8.12 (1H, d., *J* 6.9 Hz, NH); δ_C (50.31 MHz, C^2HCl_3) 71.19 (CH_2), 78.93 (CH), 133.54 & 134.05 (aromatic), 142.57 (quat. aromatic), 161.22 (urethane CO) and 176.78 (carboxyl CO).

Methyl 1-methoxy-1-amino-(N-benzyloxycarbonyl) ethanoate (91).-[349]

Compound (**90**) (11.27 g, 50 mmol) was dissolved in methanol (125 ml) and concentrated H_2SO_4 (2 ml) added. The solution was stirred at room temperature for 48 h. then poured into ice-saturated 5% $NaHCO_3$ solution (75 ml). The solution was extracted with ethyl acetate (3 x 50 ml), the organic extracts dried ($MgSO_4$) and the solvent removed under reduced pressure to give a colourless oil that solidified on treatment with light petroleum. This was purified by filtration through a short column of neutral alumina (deactivated by the addition of methanol), eluting with benzene to give the product as a colourless solid after removal of the solvent (9.87 g, 78%) of m.p. 74-76 °C (lit.[349] 76-78 °C). ν_{max} (Nujol) 3312 (NH), 1757 (ester CO), 1690 (urethane CO), 1430, 1358 & 738 cm^{-1} (aromatic); δ_H (200 MHz, C^2HCl_3) 3.43 (3H, s., $CHOCH_3$), 3.81 (3H, s., $COOCH_3$), 5.14 (2H, s., $PhCH_2O$), 5.33 (1H, d., *J* 8.2 Hz, CH), 5.90 (1H, br. d., *J* 8.2 Hz, NH) and 7.31 (5H, s., aromatic); δ_C (50.31 MHz, C^2HCl_3) 53.48 ($COOCH_3$), 56.77 (CH_3O), 67.94 (CH_2), 81.11 (CH), 128.72, 128.91 & 129.11 (aromatic), 130.66 (quat. aromatic), 156.63 (urethane CO) and 168.17 (ester CO).

Methyl 2-benzyloxycarbonylamino-2-(dimethoxyphosphinyl)-acetate (92).-[350]

Compound (**91**) (8.86 g, 35 mmol) was dissolved in toluene (35 ml) at 70 °C. Phosphorus trichloride (4.81 g, 35 mmol) was added and the solution stirred at 70 °C for 18 h. Trimethyl phosphite (4.34 g, 35 mmol) was then added and the mixture stirred at 70 °C for a further 2 h. The solvent was then removed under reduced

pressure, the residue redissolved in ethyl acetate (40 ml) and the organic layer washed with 5% $NaHCO_3$ (3 x 20 ml) and dried (Na_2SO_4). The solution was concentrated and stirred vigorously while hexane was added to precipitate the product as white crystals (7.81 g, 62%) of m.p. 78-79 °C (lit.[350] 80 °C). *m/z* (Found: $(M + H)^+$ 332.0899. $C_{13}H_{19}NO_7P$ requires 332.0899); δ_H (200 MHz, C^2HCl_3) 3.75 (3H, d., J_{PH} 7.05 Hz, 1 of $PO(OCH_3)_2$), 3.81 (3H, s., $COOCH_3$), 3.82 (3H, d., J_{PH} 7.05 Hz, 1 of $PO(OCH_3)_2$), 4.94 (1H, d.d., J_{PH} 22.4 Hz, J_{NH} 9.18 Hz, CH), 5.12 (2H, s., $PhCH_2O$), 5.80 (1H, d., J_{CH} 9.18 Hz, NH) and 7.34 (5H, s., aromatic); δ_P (121.49 MHz, C^2HCl_3) 18.74; δ_C (50.31 MHz, C^2HCl_3) 51.02 ($COOCH_3$), 53.91 (CH), 54.68 (d., J_{PC} 27.3 Hz, $PO(OCH_3)_2$), 68.10 (CH_2), 128.65-129.04 (aromatic), 136.30 (quat. aromatic), 156.06 (urethane CO) and 167.74 (ester CO); *m/z* (CI) 332 ($[M+ H]^+$, 43.4), 224 ($[M - PO(OMe)_2]^+$, 36.2), 196 ($[M - Cbz]^+$, 8.5) and 91 ($PhCH_2^+$, 100).

3.1 Molecular biology and Enzymology

Pepsin Assay (93).- Method 1

This assay is based on Knowles'[456] modifications of the method of Anson.[457] Haemoglobin (125 mg) was dissolved in water (5 ml), filtered through glass wool and the pH of the filtrate adjusted to 1.8 with 0.3 M HCl. Pepsin (1 mg) was dissolved in water (10 ml) and 10 µl of this solution added to 200 µl of the haemoglobin solution at 37 °C. After 4 min. 5% w/v trichloroacetic acid solution (400 µl) was added and the solution left at 37 °C for a further 3 min. The resulting solution was centrifuged for 3 min. and the supernatant removed carefully and its absorbance at 280 nm measured.

Method 2[292]

The peptide Cbz-His-Phe(NO_2)-Phe-OMe was dissolved in 40 mM citrate buffer (pH 4) to give concentrations in the range 0.15-0.25 mM. Typically 0.9 ml of peptide solution was added to 0.1 ml of a pepsin solution (0.007 - 0.05 mM) in the same buffer at 37 °C. The change in absorbance at 310 nm was monitored ($\Delta\varepsilon_{310}$=800).

Inhibition of pepsin with acyl hydrazides (94).-

Inhibition experiments were carried out using the hydrazides (**1**), (**2**) and (**4**). In a typical experiment a thermally equilibrated sample (0.9 ml) of a 0.2 mmol solution of the hydrazides in water was added to a 7.5 μM solution of pepsin in water (0.1 ml, pH 2). Parallel incubations were carried out, at 37 °C, and the activity of the pepsin after intervals of up to 72 h. compared to pepsin incubated under control conditions. No significant difference could be found in any of the incubations. The experiments were carried out over a range of substrate (2 - 0.01 mmol) and enzyme (7.5 - 50 μmol) concentrations. The experiments were repeated in 40 mM citrate buffer (pH 4). Similar experiments were carried out using 100 mM stocks of the hydrazides in methanol or DMF to increase the hydrazide concentration in the incubation. However, at concentrations much over 2-5 mM the hydrazides precipitated.

After 96 h. incubation the mixtures were analysed by t.l.c. in various solvent systems to see if any of the cleavage products of the hydrazides could be identified. In no case could it be unambiguously shown that any of these compounds were produced by the incubation.

Freeze-quench experiments (95).-

A typical freeze-quench experiment was carried out by adapting the method of Wang and Hofmann.[458] A solution of the quench peptide, leucine tyrosine amide (**11**), at various concentrations (2-10 mM) in 50 mM citrate buffer (0.9 ml) (pH 4.5) was added to 0.1 ml of solutions of pepsin (3-9 mg ml^{-1}) in the same buffer at 37 °C. At intervals these incubations were quenched by adding them to 20 ml of a stirred solution of sodium borohydride (5 mg ml^{-1}) in THF or 4:1 THF/ absolute ethanol at -25 °C. The solutions were then evaporated to remove the organic solvents and examined by t.l.c. on silica plates, against an authentic sample of leucinol, eluting with chloroform/ methanol (3:1). To identify the products of transpeptidation the plates were eluted with MeOH/ EtOAc/ NH_3 (35:65:3).[56a] The protein was isolated by desalting on a Sephadex G-25 column, collecting the eluate that absorbed at 280 nm and freeze-drying. The protein was examined by polyacrylamide gel electrophoresis on 12% gels. The visualisation of pepsin on these gels proved to be problematic and various staining procedures were examined.

Dephosphorylation of pepsin (96).- Method 1[298]

To pepsin (0.95 ml, 6 mg ml^{-1}) in 50 mM Tris buffer (pH 10.5 or 8.5) was added 50 or 100 µl of alkaline phosphatase solution (2.4 mg ml^{-1} in the same buffer) and the mixture incubated for 12-16 h. at 37 °C. The solutions were then examined by SDS PAGE. Further incubations were carried out at pH 6.5 (100 mM phosphate buffer). Following Perlmann's procedure,[298] incubations were most often carried out at pH 8.5.

Method 2

Pepsin (5 mg) was dissolved in water at pH 2 and pH 1, incubated overnight as above and examined by the same method.

HIV-1 protease assays (97).-

The assays are based upon the cleavage of the nonapeptide Lys-Ala-Arg-Val-Nle-(NO_2)Phe-Glu-Ala-Nle between the norleucine and nitrophenylalanine residues to give a reduction in absorbance at 300 nm. The peptide is stored as a solution in water (10 mg ml^{-1}) and the enzyme is stored as a 0.993 µM solution in 10 mM NaOAc, 0.05% 2-mercaptoethanol, 1 mM EDTA, 20% glycerol and 5% ethylene glycol. The assay buffer is;

100 mM NaOAc, 200 mM NaCl, 1 mM DTT pH 5.6.

The inhibitors were made up as 1-10 mM stock solutions in distilled MeOH and used at final concentrations in the range 1-100 µM. Methanol was added as needed to give a final concentration of 2%. Alternatively, the inhibitor was dissolved to 100 µM in assay buffer and diluted with buffer to give the required final concentration in the range 1-100 µM. For the assays the enzyme (at a final concentration of 15 nM) was added to the assay buffer and the required concentration of inhibitor and preincubated for 15 min. at 37 °C. The reaction was then initiated by the addition of 2-8 µl of substrate. The linear portion of the plot was used to give initial rates. Each inhibitor was assayed on at least two separate occasions and each concentration was tested at least twice on each occasion. The K_M for the substrate was determined

to be 33 μM under the assay conditions.

Time-dependency of inhibition (98).-

The enzyme (15 nM final concentration) and the inhibitor 3004 (see Discussion) as a 1mM stock in methanol were incubated in assay buffer at 37 °C at a final methanol concentration of 2%. Aliquots were removed and activity assayed at intervals of 6-8 min. for 30 min. However, under these conditions the activity of the native enzyme was also found to decline rapidly. Therefore, incubation at 21 °C was substituted.

Production of competent JM 101 (99).- Method 1[4]

An overnight culture of JM 101 (5 ml) was diluted into SOB medium (80 ml) and incubated at 37 °C until OD_{600} reached 0.5, when the cells were harvested by centrifugation (3,000*g*, 2 min.). The cells were resuspended in FSB buffer (40 ml) and cooled in ice for 10 min. The suspension was then centrifuged (4000*g*, 10 min.) at 4 °C, the medium decanted off and the cells resuspended in FSB buffer (4 ml). DMSO (140 μl) was then added and the cells stored on ice for 15 min., after which time a further 140 μl DMSO was added and the mixture aliquoted in 200 μl portions into sterile, chilled eppendorf tubes and frozen at -70 °C.

Method 2

An overnight culture of JM-101 (0.1 ml) was diluted into LB broth (20 ml) and incubated at 37 °C until OD_{600} reached 0.15. Aliquots (1.5 ml) were then pelleted by centrifugation (5,000*g*, 2 min.) and resuspended in ice-cold buffer A (0.5 ml). The cells were then pelleted again, resuspended in ice-cold buffer B (0.5 ml) and kept for 1 h. on ice. The cells were then frozen at -70 °C as suspensions in 15% glycerol.

Buffer A: 10 mM MOPS, 10 mM RbCl, pH 7.0

Buffer B: 100 mM MOPS, 50 mM $CaCl_2$, 10 mM RbCl, pH 6.5

PCR reaction (100).-

The oligonucleotide primers were purified by ethanol precipitation and air-dried prior to use, then made up into working solutions of 10 μM in distilled water.
The amplifications were made up as follows:

10 μl 10x Taq buffer
10 μl 2 mM dNTP's
10 μl 10 μM primer 5'-3'
10 μl 10 μM primer 3'-5'
1 μl template DNA solution
0.5 μl Taq polymerase solution
59 μl water

The mixture was covered in mineral oil (100 μl) and subjected to the following amplification cycle 25 times:

94 °C	1.5 min.	(Denaturation)
55 °C	1.5 min.	(Annealing)
72 °C	2 min.	(Extension)

The progress of the amplification was checked after the cycling process by removing a 5 μl aliquot, adding 5 μl of DNA dyes and running on an agarose gel. The gel showed one band, consistent with the presence of just one DNA sequence, the amplified one. The incubation solution was pipetted off and an equal volume of phenol added, the mixture was vortexed, centrifuged (5,000*g*, 1 min.) and the phenol removed. An equal volume of chloroform was added and the solution vortexed, centrifuged (5,000*g*, 1 min.) and the chloroform removed. To the remaining aqueous layer was added 0.1 volumes of 3 M sodium acetate solution, followed by 0.6 volumes of isopropanol and the mixture stored at -20 °C for 10 min. The precipitated DNA was pelleted by centrifugation (13,000*g*, 5 min.), the supernatant removed, the pellet washed with 70 % ethanol and air-dried.

Restriction digests (101).-

The DNA pellet from the PCR amplification was taken up in 38 µl of water and the following added:

5 µl 10x Eco R1/ Nco buffer.
5 µl 1 mg ml^{-1} BSA .
1 µl Bam H1 solution.
1 µl Eco R1 solution.

The solution was incubated at 37 °C for 4 h., after which time the DNA was purified by phenol/ chloroform extraction and ethanol precipitation as for (**100**) and taken up in water (10 µl).

In a similar fashion 5 µl of a solution of the plasmid pGEX-2T (1 mg ml^{-1}) was made up to 38 µl with water and treated exactly as for the restriction of the PCR amplified protease gene. The reaction was checked by running a 5 µl aliquot on a 1% agarose gel. The DNA was purified exactly as for the PCR restriction, to give restricted plasmid pR1/H1.

Ligation (102).-

Ligation of the restricted PCR amplified DNA into the restricted plasmid (pR1/H1) was carried out as follows:

4 µl of amplified DNA
2 µl pR1/H1
2 µl 5x ligase buffer
1 µl 4 mM ATP
0.5 µl water
0.5 µl T_4 ligase

After incubation at 37 °C for 1 to 3 hours the mixture was used directly in the tranformation step.

Transformation (103).-

The ligated plasmid (**102**) was then transfected into competent JM 101 (**99**) by addition of the crude ligation mixture (3 μl) to aliquots of competent bacteria in 200 μl buffer B. The mixture of JM 101 and ligated DNA was stored on ice for 30 min., the cells were heat-shocked at 42 °C for 90 seconds, stored on ice for a further 2 min. then 200 μl SOC medium added. After 1-2 h. at 37 °C the cells (1 ml) were plated out on LM agar containg ampicillin (100 μg ml^{-1}) at the incubation concentration and at a 1:10 dilution and left overnight at 37 °C.

Screening (104).-

Individual colonies were picked from the agar plates and grown up in 10 ml LB broth (containing 100 μg ml^{-1} ampicillin) until OD_{600} reached 0.6, when they were induced with IPTG at a final concentration of 0.1 mM. After incubation at 37 °C for a further 3 h. with shaking, 1 ml of the broth from each incubation was decanted into an eppendorf tube and centrifuged (8,000*g*, 2 min.). The supernatant was removed and the pellets resuspended in 100 μl of disruption buffer. The mixture was then boiled for 5 min. and 10 μl applied directly to a 12% SDS PAGE mini-gel. In some screens the cell pellet was sonicated before being boiled in disruption buffer for 3 min. and in others lysozyme treatment, followed by addition of Triton X-100 to 1% and vortexing, was used.

Once the gel had been run colonies that gave the largest amounts of recombinant protein were picked and grown up in 0.85 ml LB broth for 3 h., after which time glycerol (150 μl) was added and the suspensions frozen at -70 °C.

Plasmid preparation (105).-

A 500 ml culture of the transformed JM 101 was grown in LB broth containing 10% glucose to mid-log phase (OD_{600} of 0.6) and chloramphenicol added to give a 20 μg ml^{-1} solution. The culture was grown to saturation overnight, then pelleted by centrifugation (7,000*g*, 1 min.). and the cells resuspended in 50 mM glucose, 25 mM Tris.HCl pH 8.0, 10 mM EDTA (9 ml). Lysozyme (1 ml, 20 mg ml^{-1} in 200 mM Tris.

HCl, pH 7.5) was added, and the mixture left at room temperature for 5 min. A solution of 1% SDS in 200 mM sodium hydroxide (20 ml) was added, the mixture stored on ice for 5 min., followed by the addition of 5 M KOAc, (pH 5.4, 10 ml) and storage on ice for a further 15 min. The mixture was centrifuged (8,000*g*, 5 min.), the supernatant filtered through cheesecloth and 0.6 volumes of isopropanol added. After 15 min. at room temperature the solution was centrifuged (10,000*g*, 10 min.) and the supernatant removed. The pellet was washed with 70% ethanol, dried in air for 5 min. and taken up in 2 ml TE. This was split between 2 eppendorf tubes to which was added caesium chloride (1.2 g) and ethidium bromide (50 µl, 10 mg ml^{-1}), left at room temperature for 5 min. and centrifuged for 3 min. The supernatant was then removed and centrifuged (100,000*g*, 5 h.) and the plasmid band removed with a syringe. This solution was then extracted with equal volumes of butan-2-ol saturated with caesium chloride (five times) and the caesium chloride removed by dialysis against distilled water overnight. Sodium acetate (3 M, 0.1 vol) was added and the DNA precipitated by addition of 0.6 volumes of isopropanol and storage at -20 °C for 20 min. After centrifugation (10,000*g*, 5 min.) and removal of the supernatant the DNA was taken up in 0.5 ml TEN and RNase A added to a final concentration of 50 µg ml^{-1} and the mixture left at 37 °C for 30 min. The solution was then extracted with phenol (0.5 ml) and chloroform (0.5 ml) and the DNA precipitated with sodium acetate and ethanol, washed with 70% ethanol, dried and stored in distilled water at -20 °C. The optimised yield was around 200 µg from a 500 ml culture.

The BND cellulose modification was performed as follows:

The isopropanol precipitate from a 250 ml culture was taken up in 50 mM Tris.HCl, 300 mM NaCl (pH 7.6, 2 ml). This was applied to a 2 ml column of BND cellulose in the same buffer. After washing with 3 column volumes of the same buffer, the plasmid was eluted with 50 mM Tris.HCl, 1 M NaCl and 1 ml fractions collected, the plasmid precipitated with 3 volumes ethanol and treated as for the standard procedure. Plasmid was also obtained by mini-preps using the Stratagene Plasmid Quik™ kit according to the manufacturers instructions. Prior to sequencing the purity of the plasmid was checked by agarose gel electrophoresis against Eco R1/ Hind III cleaved λ phage DNA standards.

Restriction analysis of pGEX-2T (106).-

Restriction was carried out at 37 °C for 4 h. exactly as for (**101**) using the following solutions:

2 µl pGEX-2T solution
20 µl water
2.5 µl Aat II buffer
1 µl Sma1 solution
1 µl Eco RV solution

Sequencing (107).-

Sequencing of the plasmid was carried out by the dideoxy method,[417] using the USB Sequenase 2.0 kit, according to the manufacturers instructions. The primers used for the sequencing were the primers used in the PCR reaction.

Primer solution (1 µl), reaction buffer (2 µl) and template DNA (7 µl, ≈ 2 µg) were incubated at 65 °C for 2 mins., then allowed to cool slowly to room temperature in a small beaker of 65 °C water over 30 mins. The incubation was then placed on ice.

DTT (1 µl, 0.1M), was added followed by [α-^{35}S] dATP (0.5 µl, 10 µCi), labelling mix (1:8 diluted) and Sequenase™ 2.0 enzyme solution (2 µl) and the mixture incubated at room temperature for 3 mins. Labelling was terminated by addition of 3.5 µl of the incubation to 2.5 µl each of the termination mixes (A, C, G and T) and the mixtures incubated for 10 mins. Stop solution was then added to the 4 incubations, the solutions mixed thoroughly and the tubes stored on ice. The samples were heated to 75 °C for 2 mins. and then run on polyacrylamide gels (7 M in urea, pre-run for 30 mins.) at 2000 V and visualised by autoradiography on Fuji X-25 film.

Cloning into M13 (108).-

Restrictions were carried out as for (**101**) using 2 µl of M13 DNA solution or 4 µl of the PCR amplified gene.

The M13 DNA restriction was incubated at 37 °C for 2 h., then DNA dyes (10 µl)

added. The mixture was electrophoresed on a 1% SeaPlaque™ low-melting agarose gel with TAE buffer containing ethidium bromide (1 μg ml^{-1}) at 80 mA. The required band (corresponding to the cleaved RF M13) was cut from the gel and purified from the gel by the GeneClean™ kit or by melting the gel and extraction with phenol (2x), 1:1 phenol:chloroform (1x), chloroform (1x) and precipitation with 3 volumes of ethanol. The amplified gene restriction was purified as before, by phenol: chloroform extraction and ethanol precipitation or by gel electrophoresis as for the cleaved RF M13 DNA.

TAE buffer 40 mM Tris-acetate, 1 mM EDTA, pH 8.0

Ligation into M13 (109)

Ligations were carried out exactly as for (**102**), using 3.5 μl of M13 DNA solution and 2 μl of protease gene solution. The mixture was incubated at 37 °C for 1.5 h., then used directly in the transformation.

Transformation (110).-

5 μl of the ligation mixture (**109**) was added to 50 μl competent JM 101 on ice and stored for 30 min. on ice. The bacteria were then heated to 42 °C for 90 seconds, returned to ice for 2 min. and SOB without magnesium (175 μl) was added. Aliquots of the ligation mixture mixture (1, 10, 100 μl) were added to 300 μl of plating JM101 (grown up overnight in LB broth from a colony picked from an M9 agar plate). (2S)iquid SOB top agar (5 ml) containing BCIG (40 μl, 20 mg ml^{-1} in DMF) and IPTG (4 μl, 200 mg ml^{-1} in water) was added immediately. The agar was then poured onto an LB plate (prewarmed to 37 °C), spread evenly by swirling and the plates incubated at 37 °C for 12 h.

Preparation of stocks of recombinant M13 (111).-

A single clear plaque was picked from the plates above and used to inoculate 500

µl of an overnight culture of plating JM 101 in 20 ml L-broth. The culture was incubated for 4 h. and then centrifuged (12,000*g*, 5 min.) at room temperature. The supernatant was carefully dispensed into sterile 10 ml tubes and stored at 4 °C.

Plaque assay (112).-

To assess the level ot virus production from the infected cells the viral stock produced above was serially diluted ten-fold ten times. These dilutions (0.1 ml) were added to 0.1 ml of plating JM 101 in 3 ml SOB top agar and plated out onto LB agar plates. After overnight incubation the number of colonies on the plates were counted, where practical, to give an estimate of the virus density in the stock, which was found to be around 5×10^{13} pfu ml^{-1}.

Preparation of single-strand M13 DNA (113).-

Clear plaques were picked from the plates and grown up 2.5 ml of L-broth containing 25 µl of an overnight culture of JM 101 for 5 h. The cultures were then centrifuged (12,000*g*, 5 min.) at room temperature and the supernatant removed. A plaque assay of this supernatant gave a viral titre of 5×10^{13} pfu ml^{-1}.

An aliquot (2.5 ml) of this supernatant was added to 10 ml of an overnight culture of JM 101 and stored at room temperature for 5 min. The culture was used to inoculate prewarmed LB broth (250 ml) which was then incubated with shaking at 37 °C for 5 h. The culture was then centrifuged (4000*g*, 15 min.) at 4 °C, the supernatant removed and PEG 6000 (10 g) and sodium chloride (7.5 g) added with stirring. The mixture was stirred at room temperature for 45 min., then centrifuged (10,000*g*, 20 min.) at 4 °C and the supernatant removed. The solid was centrifuged again (10,000*g*, 5 min.) and the last of the supernatant removed with a drawn-out Pasteur pipette.

The pellet was dissolved in 10 mM Tris.HCl pH 8.0 (10 ml), phenol (10 ml) was added, the mixture vortexed for 2 min. and then centrifuged (3000*g*, 5 min.) The organic layer was then removed and 1:1 phenol:chloroform (10 ml) added. After vortexing and centrifugation as before the organic layer was removed and sodium

acetate (1 ml of 3 M solution) and ethanol (22 ml) were added. The solution was stored at 0 °C for 1 h. and then centrifuged (2,000*g*, 10 min.) at 4 °C. The precipitated single strand DNA was taken up in 1 ml of TE and stored at -20 °C.

STET plasmid preparations for restriction analysis (114).-

1.5 ml of an overnight culture of the transformed bacteria (**103**) was centrifuged (12,000*g*, 30 secs) and the supernatant removed. The pellet was resuspended in STET solution (350 µl) and freshly prepared lysozyme solution (10 µl, 10 mg ml^{-1} in 10 mM Tris.HCl, pH 8.0) added. The mixture was placed in a boiling water bath for 40 seconds. The lysate was centrifuged (12,000*g*, 10 min.). The bacterial debris was removed, sodium acetate (35 µl of 3 M solution) and isopropanol (420 µl) were added and the mixture stored at room temperature for 5 min. The suspension was centrifuged (12,000*g*, 5 min.) at 4 °C and the supernatant removed. The tube was dried by inverting and then 1 ml of 70% ethanol was added. The centrifugation was repeated, the supernatant removed and the tube dried in air. The pellet was dissolved in 50 µl TE containing RNAse A (5 µl of 20 µg ml^{-1}).

STET solution

0.1 M NaCl
10 mM Tris.HCl (pH 8.0)
1 mM EDTA (pH 8.0)
5% Triton X-100

5x ligase buffer

250 mM Tris.HCl pH 7.8
100 mM DTT
50 mM $MgCl_2$
5 mM spermidine

10x Taq buffer

100 mM Tris.HCl pH 8.3

500 mM KCl
15 mM $MgCl_2$
0.01% gelatin

***In vivo* assays for the protease inhibitors (115).-**

Antigen reduction assay.-

C8166 T-lymphoblastoid cells were infected with 10 $TCID_{50}$ of HIV-1 of the RF stock ($HIV\text{-}1_{RF}$) and washed three times wih PBS. Aliquots of 2 x 10^5 cells were resuspended in growth medium, RPMI 1640 supplemented by 10% fetal calf serum, with each compound to be tested at log dilutions from 0.001 to 100 μM. The cells were incubated at 37 °C for 72 h. under a 95% air/5% CO_2 atmosphere. The HIV antigen was measured in duplicate in the supernatants using a commercial ELISA assay.

Toxicity assay.-

Aliquots of 2 x 10^5 uninfected C8166 cells were suspended in 1.5 ml of the above growth medium and cultured with the test compounds at the same log dilutions for 72 h. The cells were then washed with PBS and resuspended in 200 μl of growth medium containing ^{14}C protein hydrolysate. The cells were harvested after 16 h. and the ^{14}C incorporation measured. Untreated cells were used as controls.

3.2 Molecular Modelling

3.2.1 Interactions in the S_3 site

The inhibitor MVT-101 was used as the basis for constructing a model of the interactions of our inhibitors with the protease active site. The P_4 acetyl group was removed and the side-chains at P_3-P_3' replaced with those appropriate to our inhibitor 100. The proline residue was formed by cyclisation of the P_1' norleucine

side-chain (without the terminal methyl group) onto its amide nitrogen and minimisation of the resulting ring, allowing only the ring atoms and the adjacent carbonyl groups to adjust. Minimisation was carried out using the PRCG conjugate gradient minimiser. Overlay of the ring onto the proline ring of JG-365[260] showed the two structures were very similar (r.m.s. deviation of around 0.06 Å), indicating that a reasonable ring geometry had been calculated. Minimisation was carried out in Macromodel using the AMBER forcefield[459] and the Polak-Ribiere conjugate gradient minimiser (PRCG)[460] to a convergence criterion of 0.001 kcal mol^{-1}.

The new side-chains were relaxed into their binding pockets by a substructure minimisation procedure. The substructure consisted of all residues having one or more atoms within 5Å of any atom in our inhibitor. These atoms were allowed to move freely in the minimisation. A shell of atoms around this substructure was fixed with a force constant dependent on the number of bonds between the given atom and the nearest bonded atom in the substructure. Atoms within one bond were fixed with a constant of 100 kJ mol^{-1} $Å^{-1}$, within two bonds at 500 kJ mol^{-1} $Å^{-1}$ and within 3 bonds at 1500 kJ mol^{-1} $Å^{-1}$. Any atom more than 3 bonds from the substructure was ignored in the calculation.

Once a reasonable structure had been calculated the conformational preferences of the P_3 side-chain were examined by a Monte Carlo simulation. The residues of the S_3 pocket chosen for inclusion in the calculation were those within 6Å of any atom in the P_3 residue examined:

Arg 8B, Leu 23B, Asp 29A, (Ile 47A), Gly 48A, Gly 49A, (Phe 53A) and Val 82B.

Ile 47A and Phe 53A are distant residues and probably do not contribute greatly to the binding.

The torsion angles of the side-chain were altered systematically (in 10° increments) or randomly and the energies of the resulting structures calculated. The low energy structures were then examined.

3.2.2 The configuration at phosphorus

A substructure consisting of residues 22-27 of each monomer and the central residues of the inhibitor was constructed. Molecular dynamics were carried out on

minimised structures (gradient of less than 0.01 kJ mol^{-1} $Å^{-1}$) with the Verlet algorithm.[461] A 1 fs time-step was used with SHAKE.[462] The substructure was equilibrated at 700K for 100 fs, then dynamics was carried out at 300K for 20 ps. A variant in which, after equilibration, the temperature was decreased from 700K to 300K over 10 ps, followed by 10 ps of simulation was also performed. The results were found to be similar.

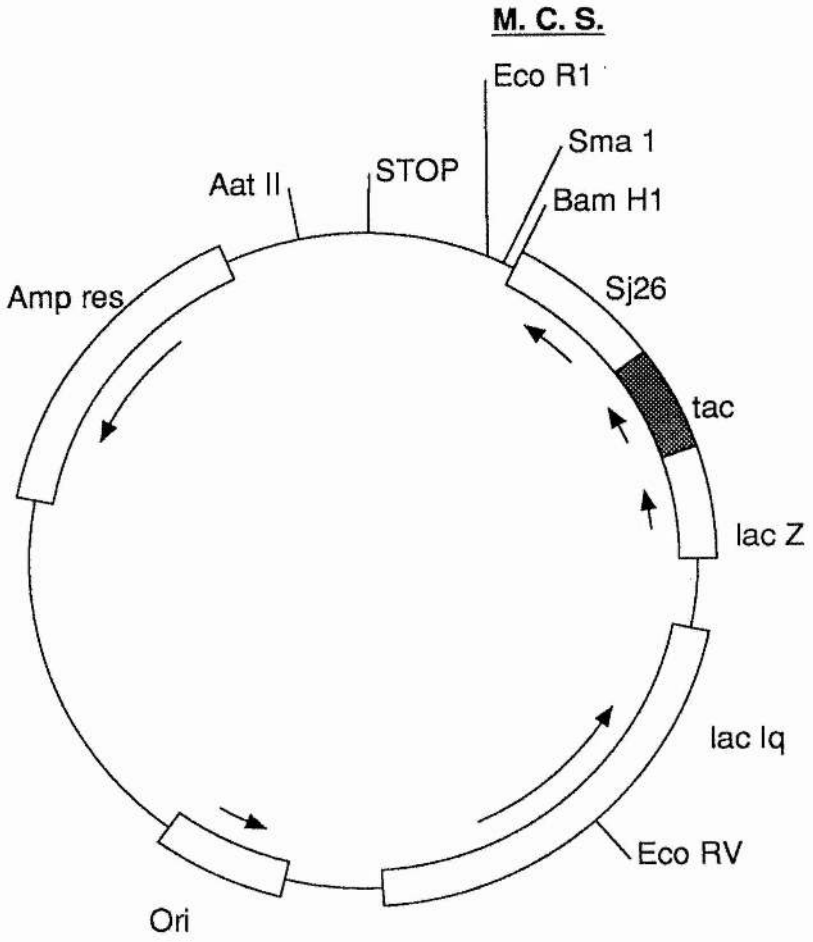
APPENDIX A

Publications

1. N. P. Camp, Paul C. D. Hawkins, P. B. Hitchcock and D. Gani, *Bioorg. Med. Chem. Lett.*, **2**, 1047 (1992).

APPENDIX B

Restriction maps of pGEX-2T and pGEX-HPR



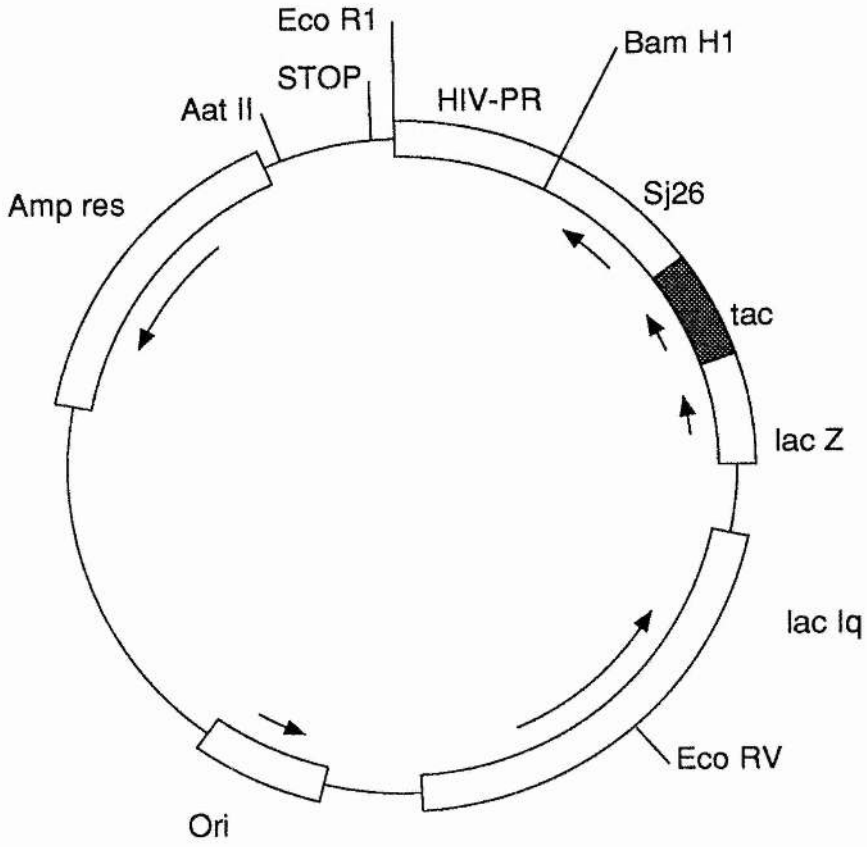
Amp res - B-lactamase
Ori - Origin of replication
lacI - lac repressor

lac Z - B-galactosidase gene
tac - tac promoter
Sj26 - glutathione transferase

M. C. S. - Multiple Cloning Site

The plasmid pGEX-2T

Plasmid pGEX-2T plus the HIV-1 protease gene



Amp res - B-lactamase
Ori - Origin of replication
lacI - lac repressor

lac Z - B-galactosidase gene
tac - tac promoter
Sj26 - glutathione transferase

4.0 References

1. H. Umezawa, T. Aoyagi, H. Morishima, M. Matzusaki, H. Hamada and T. Takeuchi, *J. Antibiot.*, **23**, 259 (1970).
2. (a) J. Drenth, *Recl. Trav. Chim. Pays-Bas*, **99**, 185 (1980); (b). M. N. G. James, *Can. J. Biochem.*, **58**, 251 (1980).
3. A. R. Fersht, *J. Am. Chem. Soc.*, **93**, 3504 (1971).
4. (a) R. F. W. Bader, J. R. Cheeseman, K. E. Laidig, K. B. Wiberg and C. M. Breneman, *J. Am. Chem. Soc.*, **112**, 6530 (1990); (b) Q.-P. Wang, A. J. Bennett, R. S. Brown and B. D. Santarsiero, *ibid.*, **113**, 5757 (1991).
5. P. Sepulveda, J. Marcziniszyn, D. Liu and J. Tang, *J. Biol. Chem.*, **250**, 5082 (1975).
6. (a) J. L. Denbury, R. Nelson and M. C. Silver, *J. Am. Chem. Soc.*, **90**, 479 (1968); (b) G. E. Clement, S. L. Snyder, H. Price and R. Cartnell, *J. Am. Chem. Soc.*, **90**, 5603 (1968); (c) T. M. Kitson and J. R. Knowles, *FEBS Lett.*, **16**, 337 (1971).
7. J. S. Fruton, *Adv. Enzymol. Relat. Areas Mol. Biol.*, **33**, 26 (1970).
8. M. N. G. James and A. R. Sielecki, *J. Mol. Biol.*, **163**, 299 (1983).
9. 'CRC Handbook of Chemistry and Physics', 59th ed., Ed. R. C. Weast, Chemical Rubber Publishing Co., Cleveland, Ohio, 1978, pg. D-202.
10. A. K. Paterson and J. R. Knowles, *Eur. J. Biochem.*, **31**, 510 (1972).
11. (a) T. G. Rajagopalan, S. Moore and W. H. Stein, *J. Biol. Chem.*, **241**, 4940 (1966); (b) R. S. Bayliss, V. R. Knowles and G. B. Wybrandt, *Biochem. J.*, **113**, 377 (1969).
12. (a) J. A. Hartsuck and J. Tang, *J. Biol. Chem.*, **247**, 2575 (1972); (b) S. C. Chen and J. Tang, *ibid.*, 2566.
13. B. F. Erlanger, S. M. Vratsanos, N. Wassermann and A. G. Cooper, *Biochem. Biophys. Res. Comm.*, **23**, 243 (1966).
14. G. R. Delpierre and J. S. Fruton, *Proc. Natl. Acad. Sci. U.S.A.*, **56**, 1817 (1966).
15. E. Gross and J. L. Morell, *J. Biol. Chem.*, **247**, 3638 (1966).
16. J. Tang, P. Sepulveda, J. Marciniszyn, K. C. S. Chen, W. Y. Huang, N. Too, D. Liu and J. P. Lapier, *Proc. Natl. Acad. Sci. U. S. A.*, **70**, 3437 (1973).
17. (a) S. K. Grant, M. L. Moore, S. A. Fakhoury, T. A. Tomaszek Jr. and T. D. Meek, *Bioorg. Med. Chem. Lett.*, **2**, 1441 (1992); (b) S. W. Kaldor, R. W. Schevitz, J.-P. Wery, D. K. Clawson, P. J. Burck, D. H. Berg, R. E. Kaiser, M. T. Lai, J. E. Munroe

and N. D. Jones, *Abstr. Papers ACS Meeting*, **203**, abstr. 178 (1992).
18. G. B. Irvine and D. T. Elmore, *Biochem. J.*, **183**, 389 (1979).
19. B. G. Rao and U. C. Singh, *J. Am. Chem. Soc.*, **113**, 6735 (1991).
20. P. S. Sampath-Kumar and J. S. Fruton, *Proc. Natl. Acad. Sci. U.S.A.*, **71**, 1070 (1974).
21. T. L. Blundell, J. A. Jenkins, B. T. Sewell, L. H. Pearl, J. B. Cooper, I. J. Tickle, B. Veerapandian and S. P. Wood, *J. Mol. Biol.*, **211**, 919 (1990).
22. J. B. Cooper, G. Khan, G. Taylor, I. J. Tickle and T. L. Blundell, *J. Mol. Biol.*, **214**, 199 (1990).
23. J. S. Richardson, *Adv. Protein Chem.*, **34**, 167 (1979).
24. T. Lah, I. Kregar, A. Sali, B. Lenorcic, M. Kotnik, V. Kostka and V. Turk, *Period. Biologorum*, **90**, 31 (1988).
25. C. Hutchins and J. Green, *CRC Crit. Rev. Biochem.*, **26**, 77 (1992).
26. X. Lin, M. Fasek, Z. Chen, G. Koelsch, H. P. Han, J. A. Hartsuck and J. Tang, 'Structure and Function of the Aspartic Proteases', Ed. B. M. Dunn, Plenum Press, New York, 1991, pg. 1.
27. (a) E. Ido, H.-P. Han, F. J. Kezdy and J. Tang, *J. Biol. Chem.*, **266**, 24359 (1992); (b) J. Tang, Y. Lin, E. Co, J. A. Hartsuck and X. Lin, *Scand. J. Clin. Lab. Invest.*, **52** (Suppl. 210), 127 (1992).
28. I. Schecter and A. Berger, *Biochem. Biophys. Res. Comm.*, **27**, 157 (1967).
29. (a) J. S. Fruton, *Adv. Enzymol. Relat. Areas Mol. Biol.*, **42**, 1 (1976); (b) T. Hofmann and R. S. Hodges, *Biochem. J.*, **203**, 603 (1982).
30. T. Hofmann, B. Allen, M. Bendiner, M. Blum and A. Cunningham, *Biochemistry*, **27**, 1140 (1988).
31. K. Teraishi, M. Saito, I. Fujii and H. Nakamura, *Tetrahedron Lett.*, **33**, 7153 (1992).
32. C.-A. Bauer, R. C. Thompson and E. R. Blout, *Biochemistry*, **15**, 296 (1976).
33. J. Lowbridge and J. S. Fruton, *J. Biol. Chem.*, **252**, 2978 (1977).
34. E. Lolis and G. Petsko, *Ann. Rev. Biochem.*, **59**, 597 (1990).
35. (a) A. J. Cornish-Bowden, P. Greenwell and J. R. Knowles, *J. Biol. Chem.*, **113**, 369 (1969); (b) J. L. Denburg, R. Nelson and M. S. Silver, *J. Am. Chem. Soc.*, **90**, 479 (1968).
36. (a) T. M. Kitson and J. R. Knowles, *Biochem. J.*, **122**, 363 (1969); (b) V. K. Antonov, I. D. Rumsh and A. G. Tikhodeeva, *FEBS Lett.*, **29**, 4 (1974); (c) N. I. Mal'tsev, L. M. Ginodman, V. N. Orekhovich, T. A. Valueva and L. N. Akimova,

Biokhimiya, **31**, 983 (1966).
37. H. Neumann, Y. Levin, A. Berger and E. Katchalski, *Biochem. J.*, **73**, 33 (1959).
38. M. S. Silver and S. L. T. James, *J. Biol. Chem.*, **255**, 555 (1980).
39. J. S. Fruton, S. Fujii and M. H. Knappenberger, *Proc. Natl. Acad. Sci. U.S.A.*, **47**, 759 (1961).
40. (a) T. M. Kitson and J. R. Knowles, *Biochem. J.*, **122**, 241 (1971); (b) *idem.*, *ibid.*, 249.
41. L. S. Shkarenkova, L. M. Ginodman, L. V. Kozlov and V. N. Orekovich, *Biokhimiya*, **33**, 154 (1968).
42. N. Sharon, V. Grisaro and H. Neumann, *Arch. Biochem. Biophys.*, **97**, 219 (1962).
43. L. V. Kozlov, L. M. Ginodman and V. N. Orekhovich, *Dok. Akad. Nauk. SSSR*, **172**, 1207 (1970).
44. L. V. Kozlov and L. L. Zavada, *Mol. Biol. (Moscow)*, **9**, 735 (1975).
45. L. M. Ginodman and N. G. Lutsenko, *Biokhimiya*, **37**, 81 (1972).
46. J. R. Knowles, *Proc. Roy. Soc. B.*, **26**, 172 (1970).
47. G. E. Clement, *Prog. Bioorg. Chem.*, **3**, 177 (1973).
48. M. S. Silver, M. Stoddard and M. H. Kelleher, *J. Am. Chem. Soc.*, **98**, 6684 (1976).
49. M. S. Silver and M. Stoddard, *Biochemistry*, **14**, 614 (1975).
50. M. W. Hunkapiller and J. H. Richards, *Biochemistry*, **11**, 2829 (1972).
51. (a) M.Takahashi and T. Hoffmann, *Biochem. J.*, **127**, 35 (1972); (b) M. Takahashi, T. T. Wang and T. Hofmann, *Biochem. Biophys. Res. Comm.*, **57**, 39 (1974); (c) T. Hoffmann, *Adv. Chem. Ser.*, **136**, 146 (1974).
52. (a) M. Akhtar and J. M. Al-Janabi, *J. Chem. Soc., Chem. Comm.*, 859; (b) *idem., ibid.*, 1002.
53. Preface to T. M. Kitson and J. R. Knowles, *J. Chem. Soc., Chem. Comm.*, 361, 1970.
54. A. K. Newmark and J. R. Knowles, *J. Am. Chem. Soc.*, **97**, 3557 (1975).
55. M. S. Silver and M. Stoddard, *Biochemistry*, **11**, 191 (1972).
56. V. K. Antonov, L. M. Ginodman, L. D. Rumsh, Y. V. Kapitannikov, T. N. Barshevskaya, L. P. Yavashev, A. G. Gurova and L. I. Volkova, *Bioorgan. Khimia*, **6**, 436 (1980).
57. M. W. Holladay, F. G. Salituro, P. G. Schmidt and D. H. Rich, *Biochem Soc. Trans.*, **13**, 1046 (1985).

58. M. Blum, A. Cunningham, H. Pang and T. Hofmann, *J. Biol. Chem.*, **266**, 9501 (1991).
59. T. Hofmann and A. L. Fink, *Biochemistry*, **23**, 5253 (1984).
60. (a) M. K. Lutek, T. Hoffmann and C. M. Deber, *J. Biol. Chem.*, **263**, 8011 (1988); (b) M. Blum, A. Cunningham, M. Bendiner and T. Hoffmann, *Biochem. Soc. Trans.*, **13**, 1044 (1985).
61. M. S. Silver and S. L. T. James, *Biochem. Biophys. Acta*, **743**, 13 (1983).
62. (a) S. W. May and E. T. Kaiser, *Biochemistry*, **11**, 592 (1972); (b) H. J. Chen and E. T. Kaiser, *J. Am. Chem. Soc.*, **96**, 1410 (1974); (c) E. Zeffren and E. T. Kaiser, *Arch. Biochem. Biophys.*, **126**, 965 (1968).
63. B. M. Dunn and A. L. Fink, *Biochemistry*, **23**, 5241 (1984).
64. T. Hofmann and R. S. Hodges, *Biochem. J.*, **203**, 603 (1982).
65. M. N. G. James, I.-N. Hsu, T. Hofmann and A. R. Sielecki, 'Structural Studies on Molecules of Biological Interest', Eds. G. Dodson, J. P. Glusker and D. Sayre, Clarendon Press, Oxford, 1981, pg. 350.
66. M. N. G. James, A. R. Sielecki, F. Salituro, D. H. Rich and T. Hofmann, *Proc. Natl. Acad. Sci. U.S.A.*, **79**, 6137 (1982).
67. J. S. Fruton, *Mol. Cell. Biochem.*, **32**, 105 (1980).
68. G. P. Sachdev and J. S. Fruton, *Proc. Natl. Acad. Sci. U.S.A.*, **72**, 3424 (1975).
69. (a) P. G. Schmidt, M. W. Holladay, F. G. Salituro and D. H. Rich, *Biochem. Biophys. Res. Comm.*, **129**, 597 (1985); (b) D. H. Rich, M. S. Bernatowicz and P. G. Schmidt, *J. Am. Chem. Soc.*, **104**, 3535 (1982).
70. (a) M. Takahashi and T. Hoffmann, *Biochem. J.*, **147**, 549 (1975); (b) M. S. Silver and M. Stoddard, *Biochemistry*, **11**, 191 (1972).
71. L. C. Spector, 'Covalent Catalysis by Enzymes', Springer Verlag, New York, 1982, pg. 9.
72. (a) K. Suguna, R. R. Bott, E. A. Padlan, E. Subramanian, S. Sheriff, G. H. Cohen and D. R. Davies, *J. Mol. Biol.*, **196**, 881 (1987); (b) K. Suguna, E. A. Padlan, C. W. Smith, W. D. Carlson and D. R. Davies, *Proc. Natl. Acad. Sci. U.S.A.*, **84**, 7009 (1987); (c) M. N. G. James and A. R. Sielecki, *Biochemistry*, **24**, 3701 (1985).
73. M. N. G. James, *Can. J. Biochem.*, **58**, 251 (1980).
74. M. Newman, M. Safro, C. Frazao, G. Khan, A. Zdanov, I. J. Tickle, T. L. Blundell and N. Andreeva, *J. Mol. Biol.*, **21**, 1295 (1991).
75. A. R. Sielecki, A. A. Fedorov, A. Boodhoo, N. S. Andreeva and M. N. G. James,

J. Mol. Biol., **214**, 143 (1990).

76. (a) A. Goldblum, *J. Molec. Struc.*, **150**, 241 (1987); (b) *idem.*, *FEBS Lett.*, **216**, 241 (1990).
77. (a) A. Goldblum, *Biochemistry*, **27**, 1653 (1988); (b) L. Turi and G. Naray-Szabo, *Int. J. Quant. Chem.*, **42**, 1537 (1992).
78. M. N. G. James and A. R. Sielecki, *Biochemistry*, **24**, 3701 (1985).
79. L. Polgar, *FEBS Lett.*, **219**, 1 (1987).
80. L. H. Pearl, *FEBS Lett.*, **214**, 8 (1987).
81. D. R. Davies, *Ann. Rev. Biochem.*, **19**, 189 (1991).
82. A. Hemmings, personal communication.
83. S. A. Bizzozero and H. Duder, *Bioorg. Chem.*, **19**, 46 (1981).
84. R. A. Eades, D. A. Well, D. A. Dixon and C. H. Douglas Jr., *J. Phys. Chem.*, **85**, 976 (1981).
85. E. M. Arnett, *Prog. Phys. Org. Chem.*, **1**, 223 (1963).
86. M. N. G. James, A. R. Sielecki, K. Hayakawa and M. H. Gelb, *Biochemistry*, **31**, 3872 (1992).
87. K. Medzihradsky, I. M. Voynick, M. Medzihradsky-Schweiger and J. S. Fruton, *Biochemistry*, **9**, 1154 (1970).
88. M. Miller, J. Schneider, B. K. Sathyanarayana, M. V. Toth, G. R. Marshall, L. Clawson, L. Selk, S. B. H. Kent and A. Wlodawer, *Science*, **246**, 1149 (1989).
89. A. Sali, B. Veerapandian, J. B. Cooper, S. I. Foundling, D. J. Hoover and T. L. Blundell, *EMBO J.*, **8**, 2179 (1989).
90. H. Slebocka-Tilk and R. S. Brown, *J. Org. Chem.*, **52**, 805 (1987).
91. V. K. Antonov, 'Aspartic Proteinases and their Inhibitors', Ed. V. Kostka, Walter de Gruyter, Berlin, 1985, pg. 203.
92. T. L. Blundell, J. Cooper, S. I. Foundling, D. M. Jones, B. Atrash and M. Szelke, *Biochemistry*, **26**, 5585 (1987).
93. K. A. Thomas, G. A. Smith, T. B. Thomas and R. S. Feldman, *Proc. Natl. Acad. Sci. U.S.A.*, **79**, 4843 (1982).
94. J. Tang and R. N. S. Wong, *J. Cell. Biol.*, **33**, 53 (1987).
95. D. J. Hoover, B. Veerapandian, J. B. Cooper, D. B. Damon, B. W. Donning, R. L. Rosati and T. L. Blundell, 'Advances in Experimental Medicine and Biology, Vol. 306: Structure and Function of the Aspartic Proteases', Ed. B. M. Dunn, Plenum Press, New York, 1991, pg. 289.
96. V. K. Antonov and S. L. Alexandrov, 'Advances in Experimental Medicine and

Biology, Vol. 306: Structure and Function of the Aspartic Proteases', Ed. B. M. Dunn, Plenum Press, New York, 1991, pg. 133.

97. A. Rayan, A. Fleiss, M. Kotler, M. Chorev and A. Goldblum, 'Advances in Experimental Medicine and Biology, Vol. 306: Structure and Function of the Aspartic Proteases', Ed. B. M. Dunn, Plenum Press, New York, 1991, pg. 375.
98. Y. Lin, M. Fusek, X. Lin, J. A. Hartsuck, F. J. Kezdy and J. Tang, *J. Biol. Chem.*, **267**, 18413 (1992).
99. S. Visser, P. J. van Rooijen, C. Schattenkerk and K. E. T. Kerling, *Biochem. Biophys. Acta,* **438**, 265 (1976).
100. A. Sali, B. Veerapandian, J. B. Cooper, D. S. Moss, T. Hofmann and T. L. Blundell, *Proteins; Struct. Funct. Genet.*, **12**, 158 (1992).
101. F. G. Salituro, N. Agarwal, T. Hofmann and D. H. Rich, *J. Med. Chem.*, **30**, 286 (1987).
102. V. Dhanaraj, C. G. Dealwis, C. Frazao, M. Badasso, B. L. Sibanda, I. J. Tickle, J. B. Cooper, H. P. C. Driessen, M. Newman, C. Aguilar, S. P. Wood, T. L. Blundell, P. M. Hobart, K. F. Geoghegan, M. J. Ammirati, D. E. Danley, B. A. O'Connor and D. J. Hoover, *Nature*, **357**, 466 (1992).
103. B. Veerapandian, J. B. Cooper, A. Sali and T. L. Blundell, *J. Mol. Biol.*, **216**, 1017 (1990).
104. K. Suguna, E. A. Padlan, R. Bott, J. Boger, K. D. Parris and D. R. Davies, *Proteins; Struct. Funct. Genet.*, **13**, 195 (1992).
105. (a) M. S. Silver and S. L. T. James, *Biochemistry*, **20**, 3183 (1981); (b) T. T. Wang and T. Hofmann, *Biochem. J.*, **153**, 701 (1976).
106. (a) G. E. Clement and S. L. Snyder, *J. Am. Chem. Soc.*, **88**, 5338 (1966); (b) T. W. Reid and D. Fahrney, *ibid.*, **89**, 3941 (1967); (c) G. E. Clement, S. L. Snyder, H. Price and R. Cartnell, *ibid.*, **90**, 5603 (1968).
107. T. R. Hollands and J. S. Fruton, *Proc. Natl. Acad. Sci. U.S.A.*, **62**, 1116 (1969).
108. A. Cunningham, M. I. Hofmann and T. Hofmann, *FEBS Lett.*, **276**, 119 (1990).
109. D. W. Green, S. Ayland, J. K. Gierse and M. E. Zupec, *Biochemistry*, **29**, 3126 (1990).
110. K. L. Rebholz and D. B. Northrop, *Biochem. Biophys. Res. Comm.*, **176**, 65 (1991).
111. K. L. Rebholz and D. B. Northrop, *FASEB J.*, **6**, A38 (1991).
112. J. D. Bernal and D. Crowfoot, *Nature*, **133**, 794 (1934).
113. M. N. G. James and A. R. Sielecki, *J. Mol. Biol.*, **163**, 299 (1983).

114. T. L. Blundell, J. A. Jenkins, B. T. Sewell, L. H. Pearl, J. B. Cooper, I. J. Tickle, B. Veerapandian and S. P. Wood, *J. Mol. Biol.*, **211**, 919 (1990).
115. (a) N. S. Andreeva, N. S. Zdanov, A. S. Gutschina and A. A. Fedorov, *J. Biol. Chem.*, **259**, 11353 (1984); (b) A. R. Sielecki, A. A. Fedorov, A. Boodhoo, N. S. Andreeva and M. N. G. James, *J. Mol. Biol.*, **214**, 143 (1990).
116. (a) M. N. G. James and A. R. Sielecki, *Nature*, **219**, 33 (1986); (b) J. A. Hartsuck, G. Koelsch and J. Remington, *Proteins: Struct. Funct. Genet.*, **13**, 1 (1992).
117. (a) M. Newman, M. Safro, C. Frazao, G. Khan, N. S. Zdanov, I. J. Tickle, T. L. Blundell and N. Andreeva, *J. Mol. Biol.*, **221**, 1295 (1991); (b) N. Andreeva, J. Dill and G. L. Gilliland, *Biochem. Biophys. Res. Comm.*, **184**, 1074 (1992).
118. (a) F. Watson, S. P. Wood, I. J. Tickle, A. Shearer, B. L. Sibanda, M. Newman, G. Khan, S. I. Foundling, J. B. Cooper, B. Veerapandian and T. L. Blundell, *Proc. 18th Linderstrom-Lang Conf.*, Elsinore, Denmark (1988); (b) M. Newman, F. Watson, P. Rowchudhury, H. Jones, M. Badasso, A. Cleasby, S. P. Wood, I. J. Tickle and T. L. Blundell, *J. Mol. Biol.*, **230**, 260 (1993).
119. (a) A. R. Sielecki, K. Hayakawa, M. Fujinaga, M. E. P. Murphy, M. Fraser, A. K. Muir, C. T. Carilli, J. A. Lewicki, J. D. Baxter and M. N. G. James, *Science*, **243**, 1346 (1989); (b) J. Rahuel, J. P. Priestle and M. G. Grütter, *J. Struc. Biol.*, **107**, 227 (1991).
120. J. B. Cooper, G. Khan, G. Taylor, I. J. Tickle and T. L. Blundell, *J. Mol. Biol.*, **214**, 199 (1990).
121. (a) N. S. Andreeva and E. Gutschina, *Biochem. Biophys. Res. Comm.*, **87**, 32 (1979); (b) T. L. Blundell, B. T. Sewell and A. D. McLachlan, *Biochem. Biophys. Acta*, **580**, 24 (1979).
122. C. Abad-Zapatero, T. J. Rydel and J. Erickson, *Proteins: Struct. Funct. Genet.*, **8**, 62 (1990).
123. A. R. Sielecki, K. Hayakawa, M. Fujinaga, M. E. P. Murphy, M. Fraser and A. K. Muir, *Science*, **243**, 1346 (1989).
124. R. Bott, E. Subramanián and D. R. Davies, *Biochemistry*, **21** 6956 (1982).
125. S. W. Fesik, J. R. Luly, J. W. Erickson and C. Abad-Zapatero, *Biochemistry*, **27**, 8297 (1987).
126. S. I. Foundling, J. Cooper, F. E. Watson, A. Cleasby, L. H. Pearl and T. L. Blundell, *Nature*, **327**, 349 (1987).
127. C. Abad-Zapatero, T. J. Rydel, D. J. Neidhart, J. Luly and J. Erickson, 'Advances in Experimental Medicine and Biology, Vol. 306: Structure and Function of the

Aspartic Proteases', Ed. B. M. Dunn, Plenum Press, New York, 1991, pg. 9.
128. B. Allen, M. Blum, A. Cunningham, G.-C. Tu and T. Hofmann, *J. Biol.Chem.*, **265**, 5060 (1990).
129. M. J. Dulton, *FEBS Lett.*, **271**, 9 (1990).
130. M. L. Bender, F. Chloupek and M. C. Neveu, *J. Am. Chem. Soc.*, **80**, 5388 (1958).
131. M. L. Bender and F. J. Kezdy, *Ann. Rev. Biochem.*, **34**, 49 (1965).
132. A. J. Kirby, R. S. McDonald and C. R. Smith, *J. Chem. Soc., Perkin Trans. II*, 1495 (1974).
133. A. J. Kirby and G. J. Lloyd, *J. Chem. Soc., Perkin Trans. II*, 1753 (1976).
134. W. P. Jencks, *Adv. Enzymol. Relat. Areas Mol. Biol.*, **43**, 219 (1975).
135. R. D. Gandour, *Bioorg. Chem.*, **10**, 169 (1981).
136. R. Kluger and J. Chin, *J. Am. Chem. Soc.*, **104**, 2891 (1982).
137. V. Somayaji, K. T. Skorey and R. S. Brown, *J. Am. Chem. Soc.*, **111**, 1445 (1989).
138. J. W. Keillor and R. S. Brown, *J. Am. Chem. Soc.*, **110**, 5205 (1991).
139. (a) V. Somayaji and R. S. Brown, *J. Am. Chem. Soc.*, **109**, 4738 (1987); (b) V. Somayaji, J. W. Keillor and R. S. Brown, *J. Am. Chem. Soc.*, **110**, 2625 (1988).
140. S. L. Aleksandrov, V. K. Antonov and P. N. Mel'nikov, *Molecular Biology (Russian)*, **18**, 1281 (1984).
141. T.C. Bruice and S. J. Benkovic, 'Bioorganic Chemistry', W. A. Benjamin, New York, 1966, Chap.1
142. V. Somayaji and R. S. Brown, *J. Org. Chem.*, **51**, 2676 (1986).
143. O. Narayan and J. E. Clements, *J. Gen. Virol.*, **70**, 1617 (1989).
144. R. F. Garry, *AIDS*, **3**, 683 (1989).
145. R. C. Gallo and L. Montaigner, *Sci. Am.*, Sept. 1990.
146. Q. J. Sattentau and R. A. Weiss, *Cell*, **52**, 631 (1988).
147. T. Shioda, J. A. Levy and C. Cheng-Meyer, *Nature*, **349**, 167 (1991).
148. T. Shioda, J. A. Levy and C. Cheng-Meyer, *Proc. Natl. Acad. Sci. U.S.A.*, **89**, 9434 (1992).
149. (a) C. Cheng-Meyer, J. T. Rutka, M. L. Rosenblum, T. McHugh, D. P. Stites and J. A.Levy, *Proc. Natl. Acad. Sci. U.S.A.*, **84**, 3526 (1987); (b) M. Tateno, F. Gonzalez-Scarano and J. A. Levy, *ibid.*, **86**, 4287 (1989).
150. L. A. Evans, T. M. McHugh, D. P. Stites and J. P. Levy, *J. Immunol.*, **138**, 3415 (1987).

151. M. I. Bukrinsky, T. L. Stanwick, M. P. Dempey and M. Stevenson, *Science*, **254**, 423 (1991).
152. B. R. Cullen, *Adv. Virus Res.*, **40**, 1 (1991).
153. T. Jacks, M. D. Power, F. R. Masiarz, P. A. Luciuw, P. J. Barr and H. E. Varmus, *Nature*, **331**, 280 (1988).
154. T. Okamoto and F. Wong-Staal, *Cell*, **47**, 29 (1986).
155. M. R. Sadaie, T. Benter and F. Wong-Staal, *Science*, **239**, 910 (1988).
156. C. A. Rosen, E. F. Terwilliger, A. I. Dayton, J. G. Sodroski and W. A. Haseltine, *Proc. Natl. Acad. Sci. U.S.A.*, **85**, 2071 (1988).
157. N. Ahmad and S. Venkatesan, *Science*, **241**, 1481 (1988).
158. K. Strebel, D. Daugherty, K. Clouse, T. Folks and M. A. Martin, *Nature*, **325**, 728 (1987).
159. A. G. Fisher, E. Collati, L. Ratner, R. C. Gallo and F. Wong-Staal, *Nature*, **326**, 216 (1988).
160. G. Nabel and D. Baltimore, *Nature*, **325**, 711 (1987).
161. M. B. Feinberger and W. C. Greene, *Curr. Opinion Immunol.*, **4**, 466 (1992).
162. M. A. Nowak, R. M. Anderson, A. R. McLion, T. F. W. Wolfs, J. Goudsmit and R. M. May, *Science*, **254**, 963 (1991).
163. (a) B. D. Preston, B. J. Poiesz and L. A. Loeb, *Nature*, **242**, 1168 (1988); (b) J. D. Roberts, K. Bebenek and T. A. Kunkel, *ibid.*, 1171.
164. J. P. Dougherty and H. M. Temin, *J. Virol.*, **62**, 2817 (1988).
165. A. G. Dalgleish, B. J. Thomason, T. C. Chanh, M. Malkovsky and R. C. Kennedy, *Lancet*, 1047 (1987).
166. A. A. Martinez, M. A. R. Marcos and A. de le Hare, *Lancet*, 454 (1988).
167. J. C. Amiesen and A. Capron, *Immunol. Today*, **12**, 102 (1991).
168. (a) H. Liu, M. A. Lampe, M. V. Iregui and H. Cantor, *Proc. Natl. Acad. Sci. U.S.A.*, **88**, 8705 (1991); (b) L. Imberti, A. Sottini, A. Bettinardi, M. Pusti and D. Primi, *Science*, **254**, 860 (1991); (c) H. M. Johnson, J. K. Russell and C. H. Pontzer, *Sci. Am.*, 42 (1992).
169. I. Roitt, 'Essential Immunology', Blackwell Scientific, 1988, pg. 129 & 149.
170. M. Lafon, M. Lafage, A. Martinez-Arends, R. Ramirez, F. Vuillen, D. Charron, V. Lotteau and D. Scott-Algara, *Nature*, **358**, 507 (1992).
171. E. F. Hounsell, D. V. Renoff, D. Liney, A. G. Dalgliesh and J. Habeshaw, *Mol. Aspects Med.*, **12**, 283 (1991).
172. H. Cantor and E. A. Boyse, *Cold Spring Harbor Symp. Quant. Biol.*, **41**, 23

(1976).
173. C. Baboonian and A. G. Dalgleish, *Mol. Aspects Med.*, **12**, 329 (1991).
174. Y. Riviere, V. Blank, P. Kourilsky and A. Israel, *Nature*, **350**, 625 (1991).
175. L. Stewart, G. Schatz and V. M. Vogt, *J. Virol.*, **64**, 5076 (1990).
176. (a) M. M. Roberts and S. Oroszlan, 'Retroviral Proteases: Control of Maturation and Morphogenesis', Ed. L. H. Pearl, Macmillan Press, 1990, pg. 131; (b) *idem.*, *Biochem. Biophys. Res. Comm.*, **160**, 486 (1989).
177. S. Oroszlan, *AIDS Res. Hum. Retroviruses*, **8**, 860 (1992).
178. F. M. Veronese, T. D. Copeland, S. Oroszlan, R. C. Gallo and M. G. Sarngadharan, *J. Virol.*, **62**, 795 (1988).
179. R. J. A. Grand, *Biochem. J.*, **258**, 625 (1989).
180. H. G. Gottinger, J. G. Sodroski and W. A. Haseltine, *Proc. Natl. Acad. Sci. U.S.A.*, **86**, 5781 (1988).
181. E. Co, J. Tang and E. Ido, *FASEB J.*, **6**, A458 (1991).
182. C. U. T. Hellen and E. Wimmer, *Experentia*, **48**, 201 (1992).
183. K. Partin, G. Zybarth, L. Ehrlich, M. DeGrombrugghe, E. Wimmer and C. Carter, *Proc. Natl. Acad. Sci. U.S.A.*, **88**, 4776 (1991).
184. H. G. Krausslich, *Proc. Natl. Acad. Sci. U.S.A.*, **88**, 3213 (1991).
185. (a) R. J. Mervis, N. Ahmad, E. P. Lillehoj, M. G. Raum, F. H. Salazar, H. W. Chan and S. Venkatesan, *J. Virol.*, **62**, 3993 (1988); (b) S. D. Gowda, B. S. Stein and E. G. Engleman, *J. Biol. Chem.*, **264**, 8459 (1989); (c) S. Erickson-Vitanen, J. Manfredi, D. E. Tribe, R. Tritch, C. A. Hutchinson III, D. D. Loeb and R. Swanstrom, *AIDS Res. Hum. Retroviruses*, **5**, 577 (1989).
186. (a) H.-G. Krausslich, R. H. Ingraham, M. T. Skoog, E. Wimmer, P. V. Pallai and C. A. Carter, *Proc. Natl. Acad. Sci. U.S.A.*, **86**, 807 (1989); (b) M. L. Moore, W. M. Bryan, S. A. Fakhoury, V. W. Magaard, W. F. Huffman, B. D. Dayton, T. D. Meek, L. Hyland, G. B. Dreyer, B. W. Metcalf, J. E. Strickler, J. G. Gorniak and C. Debouck, *Biochem. Biophys. Res. Comm.*, **159**, 420 (1989).
187. H.-G. Krausslich, *Proc. Natl. Acad. Sci. U.S.A.*, **86**, 807 (1989).
188. T. Hattori, A. Kaito, K. Takatsuki, H. Kido and N. Katunuma, *FEBS Lett.*, **248**, 48 (1989).
189. L. H. Pearl and W. R. Taylor, *Nature*, **329**, 351 (1987).
190. J. Erickson, D. J. Neihardt, J. VanDrie, D. J. Kempf, X. C. Wang, D. W. Norbeck, J.J. Plattner, J. W. Rittenhouse, M. Turon, N. Wideburg, W. E. Kohlbrenner, R. Simmer, R. Helfrich, D. A. Paul and M. Knigge, *Science*, **249**, 527 (1990).

191. S. Seelmeier, H. Schmidt, V. Turk and K. von der Helm, *Proc. Natl. Acad. Sci. U.S.A.*, **85**, 6612 (1988).
192. (a) N. E. Kohl, E. A. Emini, W. A. Scheif, L. J. Davis, J. C. Heimbach, R. A. F. Dixon, E. M. Scolnick and I. S. Sigal, *Proc. Natl. Acad. Sci. U.S.A.*, **85**, 4686 (1988); (b) S. F. L. LeGrice, J. Mills and J. Mous, *EMBO J.*, **7**, 2547 (1988).
193. R. Lapatto, T. Blundell, A. Hemmings, J. Overington, A. Wilderspin, S. Wood, J. R. Merson, P. J. Whittle, D. E. Danley, K. F. Geoghegan, S. J. Hawrylik, S. E. Lee, K. G. Scheld and P. M. Hobart, *Nature*, **342**, 299 (1989).
194. T. D. Meek, B. D. Dayton, B. W. Metcalf, G. B. Dreyer and J. E. Strickler, *Proc. Natl. Acad. Sci. U.S.A.*, **86**, 1841 (1989).
195. Y. S. Cheng, F. H. Yin, S. Foundling, D. Blomstrom and C. Kettner, *Proc. Natl. Acad. Sci. U.S.A.*, **87**, 9660 (1990).
196. P. L. Darke, C.-T. Leu, L. J. Davis, J. C. Heimbach, R. E. Diehl, W. S. Hill, R. A. F. Dixon and I. S. Sigal, *J. Biol. Chem.*, **264**, 2307 (1989)
197. G. P. Sachdev and J. S. Fruton, *Biochemistry*, **9**, 4465 (1970).
198. M. Kotler, R. A. Katz, W. Danho, J. Leis and A. M. Skalka, *Proc. Natl. Acad. Sci. U.S.A.*, **85**, 4185 (1988).
199. T. A. Tomaszek Jr., M. L. Moore, J. E. Strickler, R. L. Sanchez, J. S. Dixon, B. W. Metcalf, A. Hassell, G. B. Dreyer, I. Brooks, C. Debouck and T. D. Meek, *Biochemistry*, **31**, 10153 (1992).
200. K. Moelling, M. T. Knoop, S. Billich, I. Blaha, L. Pavlickova and M. Soucek, 'Proteases of Retroviruses', Ed. V. Kostka, Walter de Gruyter, Berlin, 1989, pg. 111.
201. J. C. Wu, S. F. Carr, K. Jarnagin, S. Kirsher, J. Barnett, J. Chow, H. W. Chan, M. S. Chen, D. Medzihradsky, D. Yamashiro and D. V. Santi, *Arch. Biochem. Biophys.*, **277**, 306 (1991).
202. P. Strop, J. Konvalinka, D. Stys, L. Pavlickova, I. Blaha, J. Velek, M. Travnick, V. Kostka and J. Sedlacek, *Biochemistry*, **30**, 3437 (1991).
203. A. G. Tomasselli, J. O. Hui, T. K. Sawyer, D. J. Staples, D. J. FitzGerald, V. K. Chaudary, I. Pastan and R. L. Heinrikson, *J. Biol. Chem.*, **265**, 408 (1990).
204. (a) P. L. Darke, C.-T. Leu, L. J. Davis, J. C. Heimbach, R. E. Diehl, W. S. Hill, R. A. F. Dixon and I. S. Sigal, *J. Biol. Chem.*, **264**, 2307 (1989); (b) M. L. Moore, W. M. Bryan, S. A. Fakhoury, V. Magaard, W. F. Huffmann, B. D. Dayton, T. D. Meek, L. Hyland, G. B. Dreyer, B. W. Metcalf, J. E. Strickler, J. G. Gorniak and C. Debouck, *Biochem. Biophys. Res. Comm.*, **159**, 420 (1989); (c) J. Toszer, A.

Gutschina, I. T. Blaha, E. M. Wondrak, S. Oroszlan, *FEBS Lett.*, **279**, 356 (1991).

205. (a) L. H. Phylip, A. D. Richards, J. Kay, J. Konvalinka, P. Strop, I. Blaha, J. Velek, V. Kostka, A. J. Ritchie, A. V. Broadhurst, W. G. Farmerie, P. E. Scarborough and B. M. Dunn, *Biochem. Biophys. Res. Comm.*, **171**, 439 (1990); (b) A. G. Tomasselli, J. O. Hui, T. K. Sawyer, D. J. Staples, C. Bannow, I. M. Reardon, W. J. Howe, D. L. DeCamp, C. S. Craik and R. L. Heinrikson, *J. Biol. Chem.*, **265**, 17675 (1990).
206. S. C. Pettit, J. Simsic, D. D. Loeb, L. Everitt, C. A. Hutchinson III and R. Swanstrom, *J. Biol. Chem.*, **266**, 14359 (1991).
207. A. Billich and G. Winkler, *Arch. Biochem. Biophys.*, **290**, 186 (1991).
208. (a) R. A. Jupp, L.-H. Phylip, J. S. Mills, S. F. J. Le Grice and J. Kay, *FEBS Lett.*, **283**, 180 (1991); (b) A. D. Richards, L.-P. Phylip, W. G. Farmerie, P. E. Scarborough, A. Alvarez, B. M. Dunn, P.-H. Hirel, J. Konvalinka, J. Strop, J. Pavlickova, V. Kostka and J. Kay, *J. Biol. Chem.*, **265**, 7733 (1990).
209. J. Richardson and D. Richardson, 'Prediction of Protein Structure and Principles of Protein Conformation', Ed. G. Fasman, Plenum Publishing Corp., New York, 1989, pg. 1.
210. (a) N. Margolis, W. Heath, E. Osborne, M. Lui and C. Vlahos, *Biochem. Biophys. Res. Comm.*, **167**, 554 (1989); (b) T. D. Meek, D. M. Lambert, B. W. Metcalf, S. R. Petteway Jr. and G. B. Dreyer, *Pharmacochem. Libr.*, **14**, 225 (1990).
211. L. E. Henderson, R. E. Benveniste, R. Sowder, T. D. Copeland, A. M. Schultz and S. Oroszlan, *J. Virol.*, **62**, 2587 (1988).
212. S. Billich, M.-T. Knoop, J. Hansen, P. Strop, J. Sedlacek, R. Mertz and K. Moelling, *J. Biol. Chem.*, **263**, 17905 (1988).
213. R. A. Poorman, A. G. Tomasselli, R. L. Heinrikson and F. J. Kezdy, *J. Biol. Chem.*, **266**, 14554 (1991).
214. R. J. Tritch, Y.-S. E. Cheng, F. H. Yin and S. Erickson-Vitanen, *J. Virol.*, **65**, 922 (1991).
215. L. J. Hyland, T. A. Tomasek, G. D. Roberts, S. A. Carr, V. W. Magaard, H. L. Bryan, S. A. Fakhoury, M. L. Moore, M. D. Minnich, J. S. Culp, R. L. DesJarlais and T. D. Meek, *Biochemistry*, **30**, 8441 (1991).
216. L. J. Hyland, T. A. Tomaszek and T. D. Meek, *Biochemistry*, **30**, 8454 (1991).
217. D. Gheysen, E. Jacobs, F. deForesta, C. Thiriart, M. Francotte, D. Thines and M.

de Wilde, *Cell*, **59**, 103 (1989).

218. R. W. Klecker Jr., J. M. Collins, R. Yarchoan, R. Thomas, J. F. Jenkins, S. Broder and C. E. Myers, *Clin. Pharmacol. Ther.*, **41**, 407 (1987).
219. G. Poli, J. M. Orenstein, A. Kinter, T. M. Kolks and A. S. Fauci, *Science*, **244**, 575 (1989).
220. D. H. Rich, J. V. N. VaraPrasad, C.-Q. Sun, J. Green, R. Mueller, K. Houseman, D. MacKenzie and M. Malkovily, *J. Med. Chem.*, **35**, 3803 (1992).
221. J. C. Craig, I. B. Duncan, D. Hockley, C. Grief, N. A. Roberts and J. S. Mills, *Antiviral Res.*, **16**, 295 (1991).
222. R. Yarchoan, R. W. Klecker Jr., K. J. Weinhold, P. D. Markham, H. K. Lyerly, D. T. Durack, E. Gelmann, S. Nuisnoff-Lehrman, R. M. Blum, D. W. Barry, G. M. Shearer, M. A. Fischl, H. Mitsuya, R. C. Gallo, J. M. Collins, D. P. Bolognesi, C. E. Myers and S. Broder, *Lancet*, 575 (1986).
223. B. A. Lardy, G. Darby and D. D. Richman, *Science*, **243**, 1731 (1989).
224. C. Debouck, *AIDS Res. Hum. Retroviruses*, **8**, 153 (1992).
225. C. A. B. Boucher, M. Tersmette, J. M. A. Lange, P. Kellam, R. E. Y. De Goede, J. W. Mulder, G. Darby, J. Goudsmit and B. A. Larder, *Lancet*, **336**, 585 (1990).
226. W. J. Greenlee, *Med. Res. Rev.*, **10**, 173 (1990).
227. R. Wolfenden, *Ann. Rev. Biophys. Bioeng.*, **5**, 271 (1976).
228. (a) M. Szelke, B. Leckie, A. Hallett, D. M. Jones, J. Suieras-Diaz, B. Atrash and A. F. Lever, *Nature*, **229**, 555 (1982); (b) M. Szelke, D. M. Jones, B. Atrash, A. Hallett and B. Leckie, 'Peptides; Structure, Function and Biology', Ed. V. J. Hruby and D. H. Rich, Pierce Chemical Co., Rockford, Illinois, 1983, pg. 579.
229. (a) P. A. Bartlett, J. E. Hanson and P. P. Giannousis, *J. Org. Chem.*, **55**, 6268 (1990); (b) G. B. Dreyer, B. W. Metcalf, T. A. Tomasek, T. J. Carr, A. C. Chandler III, L. Hyland, S. A. Fakhoury, V. M. Magaard, M. L. Moore, J. E. Strickler, C. Debouck and T. D. Meek, *Proc. Natl. Acad. Sci. U.S.A.*, **86**, 9752 (1989).
230. (a) D. H. Rich, J. Green, M. V. Toth, G. R. Marshall and S. B. H. Kent, *J. Org. Chem.*, **33**, 1285 (1990); (b) J. Erickson, D. J. Niehardt, J. VanDrie, D. J. Kempf, X. C. Wang, J. J. Plattner, J. W. Rittenhouse, M. Turon, N. Wideburg, W. E. Kohlbrenner, R. Simmer, R. Helfrich, D. A. Paul and M. Knigge, *Science*, **249**, 727 (1990); (c) N. A. Roberts, J. A. Martin, D. Kinchington, A. V. Broadhurst, J. C. Craig, I. B. Duncan, S. A. Galpin, B. K. Handa, J. Kay, A. Krohn, R. W. Lambert, J. H. Merrett, J. S. Mills, K. E. B. Parkes, S. Redshaw, A. J. Ritchie, D. L. Taylor, G. J. Thomas and P. J. Machin, *ibid*, **248**, 358 (1990); (d) T. J. Tucker, W. C. Lumma

Jr., L. S. Payne, J. M. Wai, S. J. deSolms, E. A. Giuliani, P. L. Darke, J. C. Heimbach, J. A. Zugay, W. A. Schleif, J. C. Quintero, E. A. Emini, J. R. Huff and P. S. Anderson, *J. Med. Chem.*, **35**, 2525 (1992).

231. (a) H. L. Sham, N. E. Wideburg, S. G. Spanton, W. E. Kohlbrenner, D. A. Betebenner, D. J. Kempf, D. W. Norbeck, J. J. Plattner and J. W. Erickson, *J. Chem. Soc., Chem. Comm.*, 110 (1990); (b) H. L. Sham, D. A. Betebenner, N. E. Wideburg, A. C. Saldwar, W. E. Kohlbrenner, S. Vasavonanda, D. J. Kempf, D. W. Norbeck, C. Zhao, J. J. Clement, J. E. Erickson and J. J. Plattner, *Biochem. Biophys. Res. Comm.*, **175**, 914 (1991).

232. (a) M. Miller, J. Schneider, B. K. Sathyanarayana, M. V. Toth, G. R. Marshall, L. Clawson, L. Selk, S. B. H. Kent and A. Wlodawer, *Science*, **246**, 1149 (1989); (b) J. Urban, J. Konvalinka, J. Stehlikova, E. Gregorova, P. Majer, M. Soucek, M. Andreansky, M. Falry and P. Strop, *FEBS Lett.*, **298**, 9 (1992); (c) M. Cushman, Y.-I. Oh, T. D. Copeland, S. Oroszlan and S. W. Snyder, *J. Org. Chem.*, **56**, 4161 (1991).

233. (a) T. D. Meek, D. M. Lambert, G. B. Dreyer, T. J. Carr, T. A. Tomasek, M. L. Moore, J. E. Strickler, C. Debouck, L. J. Hyland, B. W. Metcalf and S. R. Petteway, *Nature*, **343**, 90 (1990); (b) T. J. McQuade, A. G. Tomasselli, L. Liu, V. Karacostas, B. Moss, T. K. Sawyer, R. L. Heinrickson and W. G. Tarpley, *Science*, **247**, 454 (1990); (c) D. J. Kempf, D. W. Norbeck, L. Codacovi, X. C. Wang, W. E. Kohlbrenner, N. E. Wideburg, D. A. Paul, M. F. Knigge, S. Vasavanonda, A. Craig-Kennard, A. Saldwar, W. Rosenbrook, J. J. Clement, J. J. Plattner and J. Erickson, *J. Org. Chem.*, **33**, 2687 (1990); (d) S. D. Young, L. S. Payne, W. J. Thompson, N. Gaffin, T. A. Lyle, S. F. Britcher, S. L. Graham, T. H. Schultz, A. A. Deana, P. L. Darke, J. Zugay, W. A. Scheif, J. C. Quintero, E. A. Emini, P. S. Anderson and J. R. Huff, *J. Med. Chem.*, **35**, 1702 (1992); (e) M. Jaskolski, A. G. Tomaselli, T. K. Sawyer, D. G. Staples, R. L. Heinrikson, J. Schneider, S. B. H. Kent and A. Wlodawer, *Biochemistry*, **30**, 1600 (1991); (f) W. J. Thompson, P. M. D. Fitzgerald, M. K. Holloway, E. A. Emini, P. L. Darke, B. M. McKeever, W. A. Scheif, J. C. Quintero, J. A. Zugay, T. J. Tucker, J. E. Schwering, C. F. Hommick, J. Nunberg, J. P. Springer and J. R. Huff, *J. Med. Chem.*, **35**, 1685 (1992); (g) S. J. deSolms, E. Guiliani, J. P. Guare, J. P. Vacca, W. M. Sanders, S. L. Graham, J. M. Wiggins, P. L. Darke, I. S. Sigal, J. A. Zugay, E. A. Emini, W. A. Scheif, J. C. Quintero, P. S. Anderson and J. R. Huff, *J. Med. Chem*, **34**, 2852 (1991); (h) T. A. Lyle, C. M. Wiscount, J. P. Guare, W. J. Thompson, P.

S. Anderson, P. L. Darke, J. A. Zugay, E. A. Emini, W. A. Scheif, J. C. Quintero, R. A. F. Dixon, I. S. Sigal and J. R. Huff, *ibid.*, **34**, 1228 (1991); (i) J. P. Vacca, J. P. Guare, S. J. deSolms, W. M. Sanders, E. A. Guiliani, S. D. Young, P. L. Darke, J. Zugay, I. S. Sigal, W. A. Scheif, J. C. Quintero, E. A. Emini, P. S. Anderson and J. R. Huff, *ibid.*, **34**, 1225 (1991); (j) G. B. Dreyer, D. M. Lambert, T. D. Meek, T. J. Carr, T. A. Tomaszek Jr., A. V. Fernandez, H. Bartus, E. Cacciavillani, A. M. Hassell, M. Minnich, S. R. Petteway Jr., B. W. Metcalf and M. Lewis, *Biochemistry*, **31**, 6646 (1992).

234. T. F. Tam, J. Carriere, I. D. MacDonald, A. L. Castellano, D. H. Pluria, N. J. Dewdney, E. M. Thomas, C. Bach, J. Barnett, H. Chan and A. Krantz, *J. Med. Chem.*, **35**, 1318 (1992)

235. (a) P. Ashorn, T. J. McQuade, S. Thaisrivongs, A. G. Tomasselli, W. G. Tarpley and B. Moss, *Proc. Natl. Acad. Sci. U.S.A.*, **87**, 7472 (1990); (b) S. Thaisrivongs, A. G. Tomaselli, J. Hui, T. J. McQuade, S. R. Turner, J. W. Strohbach, W. J. Have, W. G. Tarpley and R. L. Heinrikson, *J. Med. Chem.*, **34**, 2344 (1991); (c) A. K. Ghosh, S. P. McKee and W. J. Thompson, *Tetrahedron Lett.*, **32**, 5729 (1991); (d) S. Thaisrivongs, A. G. Tomasselli, J. B. Moon, J. Hui, T. J. McQuade, S. R. Turner, J. W. Strobach, W. J. Howe, W. G. Tarpley and R. L. Heinrickson, *J. Med. Chem.*, **34**, 2344 (1991).

236. T. Mimoto, J. Imai. S. Kisanuki, H. Enomoto, N. Hattori, K. Akaji and Y. Kiso, *Chem. Pharm. Bull.*, **40**, 2251 (1992).

237. D. A. McLeod, R. I. Brinkworth, J. A. Ashley, K. D. Janda and P. Wirsching, *Bioorg. Med. Chem. Lett.*, **1**, 653 (1991).

238. (a) D. H. Rich, *J. Med. Chem.*, **28**, 263 (1985); (b) D. H. Rich, M. S. Bernatowicz, N. S. Agarwal, M. Kawai, F. G. Salituro and P. G. Schmidt, *Biochemistry*, **24**, 3165 (1985).

239. D. P. Dwyer and P. A. Bartlett, *Abstr. ACS Meeting*, **203**, abstr. 11, (1992).

240. (a) M. Kotler, R. A. Katz, W. Danho, J. Leis and A. M. Skalka, *Proc. Natl. Acad. Sci. U.S.A.*, **85**, 4185 (1988); (b) W. M. Bryan, S. A. Fakoury, V. M. Magaard, W. F. Huffmann, B. D. Dayton, T. D. Meek, L. Hyland, G. B. Dreyer, B. W. Metcalf, J. E. Strickler, J. G. Gorniak and C. Debouck, *Biochem. Biophys. Res. Comm.*, **159**, 420 (1989).

241. B. Stowasser, K.-H. Budt, L. Jian-Qi, A. Peyman and D. Ruppert, *Tetrahedron Lett.*, **33**, 6625 (1992).

242. R. Bone, J. P. Vacca, P. S. Anderson and M. K. Holloway, *J. Am. Chem. Soc.*,

113, 9382 (1991).

243. D. C. Humber, N. Cammack, J. A. V. Coates, K. N. Cobley, D. C. Orr, R. Storer, G. G. Weingarten and M. P. Weir, *J. Med. Chem.*, **35**, 3081 (1992).
244. R. I. Brinkworth and D. P. Fairlie, *Biochem. Biophys. Res. Comm.*, **188**, 624 (1992).
245. R. I. Brinkworth, T. C. Woon and D. P. Fairlie, *Biochem. Biophys. Res. Comm.*, **176**, 241 (1991).
246. R. I. Brinkworth, M. J. Stoermer and D. P. Fairlie, *Biochem. Biophys. Res. Comm.*, **188**, 631 (1992).
247. D. L. DeCamp, L. M. Babe, R. Salto, J. L. Lucich, M.-S. Koo, S. B. Kohl and C.S. Craik, *J. Med. Chem.*, **35**, 3426 (1992).
248. (a) J. J. Blumenstein, T. D. Copeland, S. Oroszlan and C. J. Micheda, *Biochem. Biophys. Res. Comm.*, **163**, 980 (1989); (b) K. Moelling, T. Schulze, M.-T. Knoop, J. Kay, R. Jupp, G. Nicolaou and L. H. Pearl, *FEBS Lett.*, **261**, 373 (1990).
249. T. C. Woon, R. I. Brinkworth and D. P. Fairlie, *Int. J. Biochem.*, **24**, 911 (1992).
250. (a) H. J. Schramm, H. Nakashima, W. Schramm, H. Wakayama and N. Yamamoto, *Biochem. Biophys. Res. Comm.*, **179**, 847 (1991); (b) Z.-Y. Zhang, R. A. Poorman, L. L. Maggiora, R. L. Heinrickson and F. J. Kezdy, *J. Biol. Chem.*, **266**, 15591 (1991); (c) H. J. Schramm, G. Breipohl, J. Hansen, S. Henke, E. Jaeger, C. Meichsner, G. Riess, D. Ruppert, K.-P. Rücknagel, W. Schäefer and W. Schramm, *Biochem. Biophys. Res. Comm.*, **184**, 980 (1992).
251. R. L. DesJarlais, G. I. Seibel, I. D. Kuntz, P. S. Furth, J. C. Alavarez, P. R. Ortiz de Montellano, D. L. De Camp, L. M. Babe and C. S. Craik, *Proc. Natl. Acad. Sci. U.S.A.*, **87**, 6644 (1990)
252. S. C. Tyagi, *Biochem. Cell Biol.*, **70**, 309 (1991).
253. M. A. Navia, P. M. D. Fitzgerald, B. M. McKeveer, C.-T. Leu, J. C. Heimbach, W. K. Herber, I. S. Sigal, P. L. Darke and J. P. Springer, *Nature*, **337**, 615 (1989).
254. S. Spinelli, Q. Z. Liu, P. M. Alzari, P.-H. Hirel and R. J. Poljak, *Biochimie*, **73**, 1391 (1991).
255. A. Wlodawer, M. Miller, M. Jaskolski, B. K. Sathyanarayana, E. Baldwin, I. T. Weber, L. M. Selk, L. Clawson, J. Schneider and S. B. H. Kent, *Science*, **245**, 616 (1989).
256. S. Swaminathan, W. E. Harte Jr. and D. L. Beveridge, *J. Am. Chem. Soc.*, **113**, 2717 (1991).

257. (a) W. E. Harte Jr., S. Swaminathan and D. L. Beveridge, *Proteins: Struct. Funct. Genet.*, **13**, 175 (1992); (b) D. M. York, T. A. Darden, L. G. Petersen and M. W. Anderson, *J. Am. Chem. Soc.*, **32**, 1443 (1993); (c) R. M. Venable, R. R. Brooks and F. W. Carson, *Proteins: Struct. Funct. Genet.*, **15**, 374 (1993).

258. (a) T. A. Tomaszek Jr., M. L. Moore, J. E. Strickler, R. L. Sanchez, J. S. Dixon, B. W. Metcalf, A. Hassell, G. B. Dreyer, I. Brooks, C. Debouck and T. D. Meek, *Biochemistry*, **31**, 10153 (1992); (b) K. H. M. Murthy, E. L. Winbourne, M. D. Minnich, J. S. Culp and C. Debouck, *J. Biol. Chem.*, **267**, 22770 (1992); (c) G. B. Dreyer, D. M. Lambert, T. D. Meek, T. J. Carr, T. A. Tomaszek Jr., A. V. Fernandez, H. Bartus, E. Cacciavillani, A. M. Hassell, M. Minnich, S. R. Petteway Jr., B. W. Metcalf and M. Lewis, *Biochemistry,* **31**, 6646 (1992).

259. P. M. D. Fitzgerald, B. M. McKeever, J. F. Van Middlesworth, J. P. Springer, J. C. Heimbach, C.-T. Leu, W. K. Herber and R. A. F. Dixon, *J. Biol. Chem.*, **265**, 14209 (1990).

260. A. L. Swain, M. M. Miller, J. Green, D. H. Rich, J. Schneider, S. B. H. Kent and A. Wlodawer, *Proc. Natl. Acad. Sci. U.S.A.*, **87**, 8805 (1990).

261. M. Jaskolski, A. G. Tomasselli, T. K. Sawyer, D. G. Staples, R. L. Heinrikson, J. Schneider, S. B. H. Kent and A. Wlodawer, *Biochemistry*, **30**, 1600 (1991).

262. J. Erickson, D. J. Niehardt, J. VanDrie, D. J. Kempf, X. C. Wang, D. W. Norbeck, J. J. Plattner, J. W. Rittenhouse, M. Turan, N. Wideburg, W. E. Kohlbrenner, R. Simmer, R. Helfrich, D. A. Paul and M. Knigge, *Science*, **249**, 527 (1990).

263. M. Miller, J. Schneider, B. K. Sathyanarayana, M. V. Toth, G. R. Marshall, L. Clawson, L. Selk, S. B. H. Kent and A. Wlodawer, *Science*, **246**, 1149 (1989).

264. N. Thanki, J. K. Mohana Rao, S. I. Foundling, W. J. Howe, J. B. Moon, J. O. Hui, A. G. Tomasselli, R. L. Heinrikson, S. Thaisrivongs and A. Wlodawer, *Protein Sci.*, **1**, 1061 (1992).

265. A. Wlodawer, 'Proceedings of the Robert A. Welch Foundation Conference on Chemical Research, XXXVth Conference at the Frontiers of Medicine', 1991, pg.71.

266. T. D. Meek, *J. Enzyme Inhibition*, **6**, 65 (1992).

267. T. L. Blundell, J. B. Cooper, A. Sali and Z.-Y. Zhu, 'Advances in Experimental Medicine and Biology, Vol. 306, Structure and Function of the Aspartic Proteases', Ed. B. M. Dunn, Plenum Press, New York, 1991, pg. 443.

268. L.-C. Chiang, S. Kent, R. Derango and D. Rideout, *Abstr. Papers ACS Meeting*, **203**, abstr. 17 (1992).

269. E. Fischer, *Liebig's Ann. Chem.*, **190**, 129 (1878).
270. A. Buzan, C. Egnell, P. Freon and F. Canac, *Compt. Rend. Ser. C*, **262**, 845 (1966).
271. F. Weygard and W. Steglich, *Chem. Ber.*, **92**, 3134 (1959).
272. S. S. Wang, I. D. Kuleska, D. P. Winter, R. Mahofshe, R. Kutny and J. Meienhofer, *Int. J. Pept. Protein Res.*, **11**, 297 (1978).
273. Y. K. Yur'ev, Z. V. Belyakova, P. V. Kostatki and A. I. Prokof'ev, *Zhur. Obshchei. Khim.*, **29**, 2544 (1959).
274. R. B. Kelly, *J. Org. Chem.*, **28**, 453 (1963).
275. R. B. Kelly, E. G. Daniels and J. W. Hinman, *J. Org. Chem.*, **27**, 3229 (1962).
276. S. G. Waly, *J. Chem. Soc.*, 517 (1950).
277. K. Hoffmann, A. Lindemann, M. Z. Magee and N. H. Khan, *J. Am. Chem. Soc.*, **74**, 470 (1952).
278. (a) M. Bergmann and H. Fraenhal-Conrat, *J. Biol. Chem.*, **119**, 1707 (1937); (b) *idem.*, *ibid.*, **124**, 1 (1938).
279. E. Waldschmidt-Leitz and K. Kuhn, *Chem. Ber.*, **84**, 381 (1951).
280. (a) H. B. Milne and C. Numann, *J. Am. Chem. Soc.*, **70**, 2610 (1948); (b) H. B. Milne and C. H. Peng, *ibid.*, **79**, 637 and 645 (1957); (c) H. B. Milne and C. F. Most Jr., *J. Org. Chem.*, **33**, 169 (1968).
281. R. A. Boinonas St.-Guttmann and P.-A. Jaquernod, *Helv. Chim. Acta*, **43**, 1349 (1960).
282. (a) V. Du Vigneaud and C. E. Meyer, *J. Biol. Chem.*, **98**, 143 (1932); (b) *idem*, *ibid.*, 235; (c) F. A. Levine and R. E. Steiger, *ibid.*, **93**, 581 (1931); (d) V. Du Vigneaud and R. Sealock, *ibid.*, **96**, 511 (1933); (e) U. Glocker, *Chem. Ber.*, **89**, 653 (1956); (f) J. L. O'Brien and C. Niemann, *J. Am. Chem. Soc.*, **79**, 80 (1957); (g) R. J. Kerr and C. Niemann, *J. Org. Chem.*, **23**, 893 (1958).
283. H. E Carter and C. M. Stevens, *J. Biol. Chem.*, **133**, 117 (1940).
284. A. Neuberger, *Biochem. J.*, **32**, 1452 (1938).
285. (a) V. B. Spirichev, U.-T. Chzhen, V. N. Orekhovich and L. A. Shchukina, *Biokhimiya*, **23**, 895 (1958); (b) *idem., ibid.*, **24**, 667 (1959).
286. M. Bergmann and F. Stern, *Chem. Ber.*, **63B**, 437 (1930).
287. L. Kisfaludy, T. Makacsi, M. Low and F. Drexler, *J. Org. Chem.*, **44**, 654 (1979).
288. Y. Kikugawa, K. Mitsui and T. Sakamoto, *Tetrahedron Lett.*, **31**, 243 (1990).
289. J. P. Greenstein and M. Winitz, 'Chemistry of the Amino Acids, Vol. 2', Robert E. Krieger Publishing, Florida, 1961, pg. 949

290. J. R. Knowles, H. Sharp and P. Greenwell, *Biochem. J.*, **113**, 343 (1969).
291. (a) M. L. Anson, *J. Gen. Physiol.*, **22**, 79 (1938); (b) M.L. Anson, 'Crystalline Enzymes', Eds. J. Northrup, M. Kunitz and R. M. Harriot, Columbia University Press, New York, 1948, pg. 305.
292. K. Inouye and J. S. Fruton, *Biochemistry*, **6**, 1765 (1967).
293. (a) T.-T. Wang, K. J. Dorrington and T. Hofmann, *Biochem. Biophys. Res. Comm.*, **57**, 865 (1974); (b) T.-T. Wang and T. Hofmann, *Biochem. J.*, **153**, 699 (1976); (c) *idem., ibid.*, 701.
294. V. K. Antonov, *Eur. J. Biochem.*, **117**, 195 (1981).
295. G. W. Anderson, J. E. Zimmerman and F. Callahan, *J. Am. Chem. Soc.*, **89**, 5012 (1967).
296. U. K. Laemmli, *Nature*, **227**, 680 (1970).
297. X.-L. Lin, R. N. S. Wang and J. Tang, *J. Biol. Chem.*, **264**, 4482 (1989).
298. G. E. Perlmann, *J. Am. Chem. Soc.*, **74**, 6308 (1952).
299. (a) J. Pierce and C. H. Suelter, *Anal. Biochem.*, **72**, 248 (1976); (b) V. von Klug and S. M. Hal, *ibid.*, 487; (c) J. J. Sedmak and S. E. Grossberg, *ibid.*, **79**, 544 (1977); (d) S. M. Read and D. H. Northcote, *ibid.*, **116**, 53, (1981).
300. (a) T. C. Rajagopalan, S. Moore and W. H. Stein, *J. Biol. Chem.*, **241**, 4940 (1966); (b) O. O. Blumenfeld and G. E. Perlmann, *J. Gen. Physiol.*, **42**, 553, (1953).
301. A. E. Esen, *Anal. Biochem.*, **89**, 264 (1978).
302 M. M. Bradford, *Anal. Biochem.*, **72**, 248 (1976).
303. P. G. Righetti and F. Chilemmi, *J. Chromatog.*, **157**, 243 (1978).
304. M. Akhtar and D. Gani, *Tetrahedron*, **43**, 5341 (1987).
305. V.K. Antonov, 'Aspartic Proteinases and Their Inhibitors', Ed. V. Kostka, Walter de Gruyter, Berlin, 1985, pg. 203.
306. H. Wojciechowska, B. Pawiowicz and J. Grzybowska, *Tetrahedron Lett.*, 4063 (1978).
307. Y. Gao, R. M. Hanson, J. M. Klunder, S. Y. Ko, H. Masamune and K. B. Sharpless, *J. Am. Chem. Soc.*, **109**, 5765 (1987).
308. P. H. J. Carlson, T. Katsuki, V. S. Martin and K. B. Sharpless, *J. Org. Chem.*, **46**, 3936 (1981).
309. D. Pons, M. Savignac and J.-P. Genet, *Tetrahedron Lett.*, **31**, 5023 (1990).
310. E. Taschner, L. Lubiewska, M. Smulkowski and H. Wojciechowska, *Experentia*, **24**, 521 (1968).

311. A. L. McCloskey, G. S. Forken, R. W. Kluber and W. S. Johnson, *Org. Synth., Collect. Vol. IV,* 1963, 261 and references therein.
312. P. Jouin, B. Castro, C. Zeggaf, A. Pantaloni, J. P. Senet, S. Lecolier and G. Sennyey, *Tetrahedron Lett.*, **28**, 1661 (1987).
313. L. A. Carpino and G. A. Han, *J. Org. Chem.*, **37**, 3404 (1972).
314. (a) M. Tessier, F. Albericio, A. Grandas, E. Pederoso, E. Eritja, E. Giralt, C. Granier and J. van Rietschoten, *Int. J. Pept. Protein Res.*, **22**, 125 (1983); (b) F. Albericio, A. Grandas, A. Porta, E. Pederoso and E. Giralt, *Synthesis*, 281 (1987).
315. R. L. Wolfenden and L. Frick, 'Enzyme Mechanisms', Ed. M. I. Page and A. Williams, RSC, London, 1987, pg. 97.
316. W. S. Sheldrick and M. Morr, *Acta Crystallogr., Sect. B*, **B37**, 733 (1981).
317. I. Oney and M. Caplow, *J. Am. Chem. Soc.*, **89**, 6972 (1967).
318. M. J. S. Dewar, E. G. Zoebisch, E. F. Healy and J. J. Stewart, *J. Am. Chem. Soc.*, **107**, 3902 (1985).
319. (a) R. Herranz, M. L. Suarez-Gea, S. Vinuesa, M. T. Garcia-Lopez and A. Martinez, *Tetrahedron Lett.*, **32**, 7579 (1991); (b) W. J. Moree, L. C. van Gent, G. J. van der Marel and R. M. J. Liskamp, *Tetrahedron*, **49**, 1133 (1993).
320. S. C. Pettit, J. Simsic, D. D. Loeb, L. Everitt, C. A. Hutchinson III and R. Swanstrom, *J. Biol. Chem.*, **266**, 14539 (1991).
321. V. Chavane, *Bull. Soc. Chim. Fr.*, **15**, 774 (1948).
322. K. Yanauchi, M. Kinoshita and M. Imoto, *Bull. Chem. Soc. Japan*, **45**, 2528 (1972).
323. K. Yanauchi, M. Kinoshita and M. Imoto, *Bull. Chem. Soc. Japan*, **45**, 2531 (1972)
324. M. Hariharu, S. Cheberek and A. Martell, *Synthetic Commun.*, **3**, 375 (1973).
325. E. K. Bayliss, C. D. Campbell and J. G. Dingwall, *J. Chem. Soc., Perkin Trans.1*, 2845 (1984) and references therein.
326. (a) E. K. Fields, *J. Am. Chem. Soc.*, **74**, 1528 (1952); (b) R. Tyka, *Tetrahedron Lett.*, **9**, 677 (1970).
327. K. Afarinkia, C. W. Rees and J. I. G. Cadogan, *Tetrahedron*, **46**, 7175 (1990).
328. B. Krzyzanowska and W. J. Stec, *Synthesis*, 522 (1978).
329. C. G. Gerber and D. Seebach, *Helv. Chim. Acta*, **74**, 1373 (1991).
330. D. Seebach, R, Charczuk, P. Renaud, H. Berner and H. Schneider, *Helv. Chim. Acta*, **72**, 401 (1989).

331. A. P. Gledehill, C. J. McCall and M. D. Threadgill, *J. Org. Chem.*, **51**, 3196 (1986).
332. J. Oleksyszyn, L. Subotowska and P. Mastalerz, *Synthesis*, 985 (1979).
333. P. A. Bartlett and W. B. Kezer, *J. Am. Chem. Soc.*, **106**, 4282 (1984)
334. B. Lejczak, P. Kafarski and J. Szewczyk, *Synthesis*, 412 (1982).
335. J. Oleksyszyn and L. Subotkowska, *Synthesis*, 906 (1980).
336. R. C. Corcoran and J. M. Green, *Tetrahedron Lett.*, **31**, 6827 (1990).
337. K. Omura and D. Swern, *Tetrahedron*, **34**, 1651 (1978).
338. J. R. Parikh and W. von E. Doering, *J. Am. Chem. Soc.*, **89**, 5505 (1967).
339. J. C. Collins, W. W. Hess and F. J. Frank, *Tetrahedron Lett.*, 3363 (1968).
340. G. Piancatelli, A. Scettri and M. D'Auria, *Synthesis*, 245 (1982).
341. H. C. Brown and B. C. Subba Rao, *J. Am. Chem. Soc.*, **80**, 5377 (1958).
342. (a) T. N. Sorell and R. J. Spillane, *Tetrahedron Lett.*, 2473 (1979); (b) G. W. J. Fleet and P. J. C. Harding, *ibid.*, 975 (1979); (c) T. N. Sorell and P. S. Pearlman, *J. Org. Chem.*, **45**, 3449 (1980).
343. L. I. Zakharin and I. M. Khorlina, *Tetrahedron Lett.*, 619 (1962).
344. S. Nahm and S. M. Weinreb, *Tetrahedron Lett.*, **22**, 3815 (1981).
345. W. F. Bailey and J. J. Patricia, *J. Organomet. Chem.*, **352**, 1 (1988).
346. A. McKillop, J. D. Hunt, E. C. Taylor and F. Kienzle, *Tetrahedron Lett.*, **9**, 5275 (1970).
347. M. Sting and W. Steglich, *Synthesis*, 132 (1990).
348. D. Seyferth, R. S. Marmor and P. Hilbert, *J. Org. Chem.*, **36**, 1382 (1971).
349. U. Zoller and D. Ben-Ishai, *Tetrahedron*, **31**, 863 (1975).
350. (a) U. Schmidt, A. Lieberknecht, U. Schanbacher, T. Beuttler and J. Wild, *Angew. Chem. Int. Ed. Engl.*, **21**, 776 (1982); (b) *idem., Angew. Chem. Suppl.*, 1682 (1982).
351. B. Dhawan and D. Redmore, *Phosphorus and Sulfur*, **32**, 119 (1987).
352. S. Hanessian and Y. L. Bennani, *Tetrahedron Lett.*, **31**, 6465 (1990).
353. P. A. Bartlett and K. L. McLaren, *Phosphorus and Sulfur*, **33**, 1 (1987).
354. (a) U. Schöllkopf and R. Schütze, *Liebigs Ann. Chem.*, 45 (1987); (b) A. Dehnel, J. M. Kanabus-Kaminska and G. Lavielle, *Can. J. Chem.*, **66**, 310 (1988); (c) U. Groth, L. Richter and U. Schöllkopf, *Tetrahedron*, **48**, 117 (1992); (d) S. Laschat and H. Kunz, *Synthesis*, 90 (1992); (e) F. Ouazzani, M.-L. Roumestant, P. Viallefont and A. El-Hallaoui, *Tetrahedron: Asymmetry*, **3**, 913 (1992); (f) M. Ferrari, G. Jommi, G. Miglierini, R. Pagliarin and M. Sisti, *Synth.*

Comm., **22**, 107 (1992).

355. (a) W. F. Gilmore and H. A. McBride, *J. Am. Chem. Soc.*, **94**, 4361 (1972); (b) T. Glowiak, W. Sanka-Dobrowska, J. Kowalik, P. Mastalerz, M. Soroka and J. Zon, *Tetrahedron Lett*, 3965 (1977); (c) A. Kotynski and W. J. Stec, *J. Chem. Res.*, 41 (1978).
356. R. Huber, A. Knierzinger, J.-P. Obrecht and A. Vasella, *Helv. Chim. Acta*, **68**, 1730 (1985).
357. S. Shatzmiller, B.-Z. Dolitzky, R. Meirovich, R. Neidlen and C. Weik, *Liebigs Ann. Chem.*, 161 (1991).
358. A. Vasella and R. Voeffray, *Helv. Chim. Acta*, **65**, 1953 (1982).
359. T. Schrader, R. Kober and W. Steglich, *Synthesis*, 372 (1986).
360. M. Sawamura, Y. Ito and T. Hayashi, *Tetrahedron Lett.*, **30**, 2247 (1989).
361. S. E. Denmark, N. Chatani and S. V. Pansare, *Tetrahedron*, **48**, 2191 (1992).
362. U. Schollkopf, I. Hoppe and A. Thiele, *Liebigs Ann. Chem.*, 555 (1985).
363. T. Yokomatsu and S. Shibuya, *Tetrahedron: Asymmetry*, **3**, 377 (1992).
364. J. Kowalik, W. Sanka-Dobrowolska and T. Glowiak, *J. Chem. Soc., Chem. Comm.*, 446 (1984).
365. (a) G. Lavielle, P. Hautefauve, C. Schaeffer, J. A. Boutin, C. A. Cudennec and A. Pierre, *J. Med. Chem.*, **34**, 1998 (1991); (b) L. Maier and H. Spörri, *Phosphorus, Sulfur and Silicon*, **61**, 69 (1991).
366. P. Kafarski, B. Lejczak and J. Szewczyk, *Can. J. Chem.*, **61**, 2425 (1991).
367. V. A. Solodenko, T. N. Kasheva, V. P. Kukar, E. V. Kozlova, D. A. Mironenko, V. K. Svedas and A. N. Belozersky, *Tetrahedron*, **47**, 3989 (1991).
368. U. Schöllkopf and R. Schütze, *Liebigs Ann. Chem.*, 45 (1987)
369. M. Soroka and J. Zygmunt, *Synthesis*, 370 (1988).
370. S. J. Field and D. W. Young, *J. Chem. Soc., Perkin Trans. 1*, 591 (1982).
371. J. Oleksyszyn and J. C. Powers, *Biochemistry*, **30**, 485 (1991).
372. D. E. Tronrud, H. M. Holden and B. W. Matthews, *Science*, **235**, 571 (1987).
373. (a) C.-M. Kam, N. Nishino and J. C. Powers, *Biochemistry*, **18**, 3032 (1979); (b) J. E. Hanson, A. P. Kaplan and P. A. Bartlett, *ibid.*, **28**, 6294 (1989).
374. P. A. Bartlett, J. E. Hanson and P. P. Giannousis, *J. Org. Chem.*, **55**, 6268 (1990).
375. P. A. Bartlett, C. K. Marlowe, P. P. Giannousis and J. E. Hanson, *Cold Spring Harbor Symp. Quant. Biol.*, **52**, 83 (1987).
376. D. Grobelny, E. M. Wondrak, R. E. Galardy and S. Oroszlan, *Biochem.*

Biophys. Res. Comm., **169**, 1111 (1990).

377. G. B. Dreyer, B. W. Metcalf, T. A. Tomaszek Jr., A. C. Chandler III, L. Hyland, S. A. Fakhoury, V. W. Magaard, M. L. Moore, J. E. Strickler, C. Debouck and T. D. Meek, *Proc. Natl. Acad. Sci. U.S.A.*, **86**, 9752 (1989).
378. S. Ikeda, J. A. Ashley, P. Wirsching and K. D. Janda, *J. Am. Chem. Soc.*, **114**, 7604 (1992).
379. A. Peyman, K.-H. Budt, J. Spanig, B. Stowasser and D. Ruppert, *Tetrahedron Lett.*, **33**, 4549 (1992).
380. P. A. Bartlett, J. E. Hanson and P. P. Giannousis, *J. Org. Chem.*, **55**, 6268 (1990).
381. P. A. Bartlett and F. Acher, *Bull. Soc. Chim. Fr.*, 771 (1986).
382. M. C. Allen, W. Furher, B. Tuck, R. Wade and M. J. Wood, *J. Med. Chem.*, **32**, 1652 (1989).
383. B. Stowasser, K.-H. Budt, L. Jian-Qi, A. Peyman and D. Ruppert, *Tetrahedron Lett.*, **33**, 6625 (1992).
384. G. B. Dreyer, J.-K. Choi, T. D. Meek and T. A. Tomaszek Jr., *Abstr. Papers ACS Meeting*, **203**, abst. 179 (1992).
385. D. A. McLeod, R. I. Brinkworth, J. A. Ashley, K. D. Janda and P. Wirsching, *Bioorg. Med. Chem. Lett.*, **1**, 653 (1991).
386. N. I. Tarasova, S. V. Gulnik, A. A. Prischenko, M. V. Livontsov, E. N. Lysogorskaya and E. S. Oksenoit, 'Advances in Experimental Medicine and Biology, Vol. 306: Structure and Function of the Aspartic Proteases', Ed. B. M. Dunn, Plenum Press, New York, 1991, pg. 287.
387. J. T. Griffiths, L. H. Phylip, J. Konvalinka, P. Strop, A. Gutschina, A. Wlodawer, R. J. Davenport, R. Briggs, B. M. Dunn and J. Kay, *Biochemistry*, **31**, 5193 (1992).
388. S. P. Jordan, J. Zugay, P. L. Darke and L. C. Kuo, *J. Biol. Chem.*, **267**, 20028 (1992).
389. M. Dixon, *Biochem J.*, **55**, 70 (1953).
390. M. Cushman, Y.-I. Oh, T. D. Copeland, S. Oroszlan and S. W. Snyder, *J. Org. Chem.*, **56**, 4161 (1991).
391. (a) J. S. Fruton in 'Peptides 1972', Eds. H. Hanson and H. D. Jakubke, North-Holland, Amsterdam, 1973, pg. 355; (b) J. R. Knowles, H. Sharp and P. Greenwell, *Biochem. J*, **113**, 343 (1969).
392. J. T. Griffiths, L. H. Phylip, J. Konvalinka, P. Strop, A. Gutschina, A. Wlodawer, R. J. Davenport, R. Briggs, B. M. Dunn and J. Kay, *Biochemistry*, **31**, 5193 (1992).

393. (a) C. Debouck, *AIDS Res. Hum. Retroviruses*, **8**, 153 (1992); (b) S. Kageyama, J. H. Weinstein, T. Shirasaka, D. J. Kempf, D. W. Norbeck, J. J. Plattner, J. Erickson and H. Mitsuya, *Antimicrob. Agents Chemotherapy*, **36**, 926 (1992).
394. J. Maple, U. Dinur and A. T. Hagler, *Proc. Natl. Acad. Sci. U.S.A.*, **85**, 5350 (1988).
395. J. Maple, T. S. Thatcher, U. Dinur and A. T. Hagler, *Chem. Design Automation News*, **5**, 5 (1990).
396. I. T. Weber, M. Miller, M. Jaskolski, J. Leis, A. M. Skalka and A. Wlodawer, *Science*, **243**, 928 (1989).
397. (a) D. H. Rich and E. T. O. Sun, *Biochem. Biophys. Res. Comm.*, **27**, 157 (1980); (b) P. A. Bartlett and W. B. Kezer, *J. Am. Chem. Soc.*, **106**, 4282 (1984).
398. (a) H. M. Holden, D. E. Tronrud, A. F. Monzingo, L. H. Weaver and B. W. Matthews, *Biochemistry*, **26**, 8542 (1987); (b) P. A. Bartlett and C. K. Marlowe, *ibid.*, 8553.
399. D. H. Rich and M. S. Bernatowicz, *J. Med. Chem.*, **25**, 791 (1982).
400. F. Cumin, D. Nisato, J. P. Gagnol and P. Corvol, *Biochemistry*, **26**, 7615 (1987).
401. W. M. Kati, D. T. Pals and S. Thaisrivongs, *Biochemistry*, **26**, 7621 (1987).
402. T. S. Angeles, G. A. Roberts, S. A. Carr and T. D. Meek, *Biochemistry*, **31**, 11778 (1992).
403. L. Frick, R. V. Wolfenden, E. Smal and D. C. Baker, *Biochemistry*, **25**, 1616 (1986).
404. (a) A. G. Tomasselli, M. K. Olsen, J. O. Hui, D. J. Staples, T. K. Sawyer, R. L. Heinrickson and C.-S. C. Tomich, *Biochemistry*, **29**, 264 (1990); (b) P. L. Darke, C.-T. Leu, L. J. Davis, J. C. Heimbach, R. E. Diehl, W. S. Hill, R. A. F. Dixon and I. S. Sigal, *J. Biol. Chem.*, **264**, 2307 (1989); (c) M. C. Graves, J. J. Lim, E. P. Heimer and R. A. Kramer, *Proc. Natl. Acad. Sci. U.S.A.*, **85**, 2449 (1988); (d) J. E. Strickler, J. Gorniak, B. Dayton, T. D. Meek, M. Moore, V. Magaard, C. Malinowski and C. Debouck, *Proteins: Struct. Funct. Genet.*, **6**, 139, (1989); (e) B. D. Korant and C. J. Rizzo, *Biol. Chem. Hoppe-Seyler*, **371**, 271 (1990); (f) S. Erickson-Vitanen, *AIDS Res. Hum. Retroviruses*, **5**, 577 (1989); (g) T. D. Meek, B. D. Dayton, B. W. Metcalf, G.B. Dreyer, J. E. Strickler, J. G. Gorniak, M. Rosenberg, M. L. Moore, V. W. Magaard and Debouck, *Proc. Natl. Acad. Sci. U.S.A.*, **86**, 1841 (1989); (h) C.-Z. Giam and I. Boros, *J. Biol. Chem.*, **263**, 14617 (1988); (i) D. E. Danley, K. F. Geoghegan, K. G. Scheld, S. E. Lee, J. R. Merson, S. J. Hawrylik, G. A. Rickett, M. J. Ammirati and P. M. Hobart, *Biochem.*

Biophys. Res. Comm., **165**, 1043 (1989); (j) J. Boutelje, A. R. Karlström, M. G. N. Hartmanis, E. Holmgren, A. Sjögren and R. L. Levine, *Arch. Biochem. Biophys.*, **283**, 141 (1989).

405. H. A. Overton, Y. Fuji, I. R. Price and I. M. Jones, *Virology*, **170**, 116 (1989).
406. S. Pichuantes, L. M. Babe, P. J. Barr and C. S. Craik, *Proteins: Struct. Funct. Genet.*, **6**, 324 (1989).
407. (a) T. D. Copeland and S. Oroszlan, *Gene. Anal. Tech.*, **5**, 109 (1988); (b) R. F. Nutt, S. F. Brady, P. L. Darke, T. M. Ciccarone, C. D. Colton, E. M. Nutt, J. A. Rodkey, C. D. Bennet, L. H. Waxman, I. S. Sigal, P. S. Anderson and D. F. Veber, *Proc. Natl. Acad. Sci. U.S.A.*, **85**, 7129 (1988); (c) M. Schnölzer and S. B. H. Kent, *Science*, **256**, 221 (1992); (d) R. D. de L. Milton, S. C. F. Milton and S. B. H. Kent, *ibid.*, 1445.
408. D. B. Smith and K. S. Johnson, *Gene*, **67**, 31 (1988).
409. J. Messing, *Recomb. DNA Tech. Bull.*, **2**, 43 (1979).
410. S. L. Schreiber and G. L. Verdine, *Tetrahedron*, **47**, 2543 (1991).
411. (a) R. K. Saiki, S. Scharf, F. Faloona, K. B. Mullis, G. T. Horn, H. A. Ehrlich and N. Arnheim, *Science*, **230**, 1350 (1985); (b) K. B. Mullis and F. Faloona, *Methods Enzymol.*, **155**, 335 (1987).
412. D. B. Smith, K. M. Davern, P. G. Board, W. U. Tiu, E. G. Garcia and G. F. Mitchell, *Proc. Natl. Acad. Sci. U.S.A.*, **83**, 8703 (1986).
413. H. A. De Boer, L. J. Comstock and M. Vasser, *Proc. Natl. Acad. Sci. U.S.A.*, **80**, 21 (1983).
414. D. P. Gearing, N. A. Nicola, D. Metcalfe, T. A. Willson, N. M. Gough and R. L. Williams, *Bio/Technology*, **7**, 1157 (1989).
415. D. S. Montgomery, O. M. P. Singh, N. M. Gray, C. W. Dykes, M. P. Weir and A. N. Hobden, *Biochem. Biophys. Res. Comm.*, **175**, 785 (1991).
416. D. Hanahan, *J. Mol. Biol.*, **166**, 557 (1983).
417. (a) F. Sanger, S. Nicklen and A. R. Coulson, *Proc. Natl. Acad. Sci. U.S.A.*, **74**, 5463 (1977); (b) R. B. Wallace, M. J. Johnson, S. Y. Suggs, K. Miyoshi, R. Bhatt and K. Itakura, *Gene*, **16**, 21 (1981).
418. J. Rittenhouse, M. C. Turon, R. J. Helfrich, K. S. Albrecht, D. Weigl, R. L. Simmer, F. Mordini, J. Erickson and W. E. Kohlbrenner, *Biochem. Biophys. Res. Comm.*, **171**, 60 (1990).
419. J. M. Louis, R. A. McDonald, N. T. Nashed, E. M. Wondrak, D. M. Jerina, S. Oroszlan and P. T. Mora, *Eur. J. Biochem.*, **199**, 361 (1991).

420. M. Kotler, G. Arad and S. H. Hughes, *J. Virol.*, **66**, 6781 (1992).
421. (a) J. Hartmann, P. Doram, R. A. Frissell, T. Rado, D. J. Beng and E. J. Sorscher, *Biotech. Bioeng.*, **39**, 828 (1992); (b) C. H. Schein, *Bio/Technol.*, **7**, 1141 (1989).
422. R. Deveruex, G. Haeberli and M. F. Smithies, *Nuc. Acids Res.*, **12**, 387 (1984).
423. X. Ji, P. Zhang, R. N. Armstrong and G. L. Gilliland, *Biochemistry*, **31**, 10169 (1992).
424. D. H. Rich and E. T. O. Sun, *Biochem. Pharmacol.*, **29**, 2205 (1980).
425. B. Corr and A. Wilderspin, *Biochem. Soc. Trans.*, **20**, 160 (1992).
426. I.U.P.A.C., 'Nomenclature of Organic Chemsitry', Pergamon Press, Oxford 1979.
427. W. C. Still, M. Kahn and A. Mitra, *J. Org. Chem.*, **43**, 2923 (1978).
428. D. D. Perrin and W. L. F. Amarego, 'Purification of Laboratory Chemicals', Pergamon Press, Oxford, 1988.
429. J. Sambrook, E. F. Fritsch and T. Maniatis, 'Laboratory Manual of Molecular Cloning', 2nd Ed., Cold Spring Harbour Laboratory Press, Cold Spring Harbour, 1989 .
430. F. Mohamadi, N. G. J. Richards, W. C. Guida, R. Liskamp, M. Lipton, C. Caufield, G. Chang, T. Hendrickson and W. C. Still, *J. Comp. Chem.*, **11**, 440 (1990).
431. BIOSYM Technologies, Unit 17, Building 2, INTEC Technology Centre, Wade Road, Basingstoke, Hampshire.
432. F. C. Bernstein, T. F. Koetzle, G. J. B. Williams, E. F. Meyer Jr., M. D. Brice, J. R. Rodgers, O. Kennard, M. Shizmanouchi and M. Tasumi, *J. Mol. Biol.*, **112**, 535 (1977).
433. R. A. Boissonnas-St. Guttmann, P. A. Jaquernod and J. R. Waller, *Helv. Chim. Acta*, **38**, 1491 (1955).
434. *Beil.* **14**, 612.
435. S. Klimkiewicz and S. Drobarek, *Rocz. Chem.*, **49**, 959 (1973).
436. *Beil.*, **14** (2nd supp.), 298.
437. H. Schwarz, F. M. Bumpus and I. H. Page, *J. Am. Chem. Soc.*, **7**, 5697 (1957).
438. B. Weinstein and A. E. Pritchard, *J. Chem. Soc. (C)*, 1015 (1972).
439. G. Kokotes, *Synthesis*, 1990, 299.
440. R. Schwyzer, B. Iselin, H. Kappeler, B. Rinther, W. Ritter and H. Zuber, *Helv. Chim. Acta*, **41**, 1273 (1958).
441. *Beil.* **14**, 298.

442. *Beil.*, **4**, 518.
443. M. Bodansky and A. Bodansky, 'The Practice of Peptide Synthesis', Springer-Verlag, New York, 1984, pg.24.
444. G. W. Anderson, J. E. Zimmerman and F. M. Callahan, *J. Am. Chem. Soc.*, **89**, 5012 (1967).
445. D. E. Nitecki, B. Halpern and J. W. Westley, *J. Org. Chem.*, **33**, 864 (1968).
446. (a) B. Iselin, M. Feurer and R. Schwyzer, *Helv. Chim. Acta*, **38**, 1508 (1955); (b) M. Bodansky and A. Bodansky, 'The Practice of Peptide Synthesis', Springer-Verlag, New York, 1984, pg.177.
447. E. L. Smith, D. H. Spackman and W. J. Polglase, *J. Biol. Chem.*, **199**, 801 (1952).
448. G. M. Bonora, A. Maglione and C. Toniolo, *Polymer*, **15**, 767 (1974).
449. O. Keller, W. E. Keller, G. van Look and G. Wersin, *Org. Syn.*, 99 (1980).
450. G. W. Anderson and A. C. McGregor, *J. Am. Chem. Soc.*, **79**, 6180 (1957).
451. P. J. Smith and A. N. Bournes, *Can J. Chem.*, **52**, 749 (1974).
452. (a) T. N. Sorell and R. J. Spillane, *Tetrahedron Lett.*, 2473 (1979); (b) G. W. J. Fleet and P. J. C. Harding, *ibid.*, 975 (1979); (c) T. N. Sorell and P. S. Pearlman, *J. Org. Chem.*, **45**, 3449 (1980).
453. E. J. Corey and J. W. Suggs, *Tetrahedron Lett.*, 2647 (1975).
454. R. B. Merrifield, *J. Biol. Chem.*, **232**, 43 (1958).
455. H. A. Barker, R. D. Smyth, E. J. Wawszkiewics, M. N. Lee and R. M. Wilson, *Arch. Biochem. Biophys.*, **78**, 468 (1958)
456. J. R. Knowles, H. Sharp and P. Greenwell, *Biochem. J.*, **113**, 343 (1969).
457. (a) M. L. Anson, *J. Gen. Physiol.*, **22**, 79 (1938); (b) M.L. Anson, 'Crystalline Enzymes', Eds. J. Northrup, M. Kunitz and R. M. Harriot, Columbia University Press, New York, 1948, pg. 305.
458. (a) T.-T. Wang, K. J. Dorrington and T. Hofmann, *Biochem. Biophys. Res. Comm.* **57**, 865 (1975); (b) T.-T. Wang and T. Hofmann, *Biochem. J.*, **153**, 699 (1976); (c) *idem., ibid.*, 701.
459. (a) S. J. Weiner, P. A. Kollman, D. Case, U. C. Singh, G. Alagona, S. Profeta and P. Weiner, *J. Am. Chem. Soc.*, **106**, 765 (1984); (b) S. J. Weiner, P. A. Kollman, D. T. Nguyen and D. Case, *J. Comp. Chem.*, **7**, 230 (1986).
460. E. Polak and G. Ribiere, *Revue Francais Informat. Recherches Operationelle*, **16**, 35 (1969).
461. L. Verlet, *Phys. Rev.*, **159**, 98 (1967).

462. (a) W. F. Van Gunsteren and H. J. C. Berendsen, *Mol. Phys.*, **34**, 1311 (1977); (b) J. P. Ryckaert, G. Ciccotti and H. J. C. Berendsen, *J. Comp. Phys.*, **23**, 237 (1977).